

# One veil less for Japanese science

Will the Japanese science bureaucracy's drive to end oriental inscrutability serve simply to whet Western appetites for more information?

OBSERVANT readers may have noticed that news of the Japanese government's budget requests for science and technology (see page 783) are being published in *Nature* a week or two earlier than usual. That may not seem a remarkable change but it is in fact an indicator of new and welcome efforts by several of Japan's science-related ministries and agencies to communicate more effectively with the outside world.

Budget requests are normally released to the Japanese press at the end of August and until that time government officials would never grant interviews on the requests of the foreign press (unless they had some pet project they wanted to promote). But last Friday, nearly a week before the budget requests were to be officially submitted to the Ministry of Finance, the Agency of Industrial Science and Technology (AIST) of the Ministry of International Trade and Industry (MITI) put on a seven-hour series of lectures by 13 heads of department on AIST's budget requests. A similar meeting was held by the Science and Technology Agency (STA) on Monday. The language of the meetings was primarily English, which was a mark of the generosity of the Japanese participants, although that matters less than that the meetings took place. For anyone familiar with the well-oiled public relations operations of government in Washington, DC, for example, this may not seem particularly unusual. But in Japan such willingness to open the machinery of government to foreigners is unprecedented.

MITI and STA may be rightfully proud of what they have done. Other of Japan's science-related ministries will now be put in a position of having to follow suit or seeming hopelessly parochial. Even the highly conservative Ministry of Education, Science and Culture (MESC) shows signs of welcoming foreign interest. Earlier this month the ministry proudly presented the foreign press with descriptions in English of the ministry's awards of "special distinguished" grants — a welcome move but one nevertheless long overdue given that the grants are intended to be for "internationally recognized research."

The reasons for the Japanese government's interest in the foreign science press are not hard to fathom. In the case of MITI, some people within the ministry have argued for the past several years that all of its major research projects should be opened to participation by foreign researchers and by foreign companies. Once this became the consensus view and passed into official policy

in June, MITI officials realized that if they did not explain their projects to the Western world then their new policy would have been established in vain.

The more interesting question is what will happen next. The new policy strips away one of the veils that surround decision-making in the Japanese government and will help the Western world to receive some early warning of new initiatives in science and technology. But in important ways, Japan remains quite different from its chief competitor in the West.

In the United States, numerous organizations that are independent of government help to assess and decide national policy, among the more influential being the National Academy of Sciences and the congressional Office of Technology Assessment. The results of their deliberations are prepared for public distribution. These relatively transparent organizations have no equivalents in Japan, where enigma still surrounds the deliberations which gives birth to a government budget request. MITI, STA (and other organizations which jump on their bandwagon) must expect that having seen one veil removed, foreigners will hope that others will soon follow. □

## Reactor woes

Berlin's Hahn-Meitner Institute is in danger. West German researchers could be doing more to save it.

RESEARCHERS in West Germany should stand up for the embattled Hahn-Meitner Institute, now in danger of being shut down as a result of the West Berlin city government's refusal to license a newly renovated reactor upon which the institute depends for its research. The Hahn-Meitner Institute, which is one of 13 national laboratories, may not be able to survive if the reactor is not licensed soon (see *Nature* 346, 688).

On the face of it, the battle for the Hahn-Meitner Institute has little to do with research and much to do with the antipathy to nuclear power that is prevalent in West Germany. Opponents of nuclear power in West Berlin, which actually has no nuclear-power reactors of its own, have simply latched onto the institute as a convenient target for a crusade against nuclear energy. Their argument — that a government must offer an adequate storage site for nuclear waste *before* it grants an operating



license to any reactor — is a rational one. But it makes no sense when its prime object becomes a research reactor, which will generate only a few kilogram of waste every year. Such reasoning is like “shooting at sparrows with cannons”, as West Berlin mayor Walter Momper said recently (before he too lined up with those who oppose the reactor).

Apart from a weakly-worded letter from the association of national laboratories (AGF), the response from the research community has been nil. Researchers in other national laboratories and at universities could and should be making a broader and more concerted effort to save the Hahn-Meitner Institute — by petitioning politicians, for instance, or mounting a publicity campaign emphasizing the importance of the reactor, which will serve solely as a source of neutrons, for basic science.

Perhaps researchers in West Germany are afraid to be seen defending any institution linked to the word “nuclear”. Perhaps they themselves are opposed to licensing the reactor. But a more likely reason why researchers did not defend their colleagues at the Hahn-Meitner Institute lies in the zero-sum view of research spending. In this way of thinking, closure of one national laboratory means more money for the rest to share.

That is a view that is likely to be disappointed. There is no shortage of politicians and voters in Germany who are opposed not just to nuclear energy but to science *per se* and who would welcome the chance to roll back at least part of the tremendous post-war expansion of basic and applied research. For them, the Hahn-Meitner Institute is only one of many unnecessary research enterprises. West German scientists should realize that the closure of the institute will not mean more for those that remain but, ultimately, less for all. □

## Big Bang hypothesis

A new section of *Nature* is not a vehicle for the publication of half-baked ideas.

FAITHFUL readers should not be too mystified by the appearance, on page 807 of this issue, of a newly named section of this journal. The rubric HYPOTHESIS is intended as an occasional vehicle for scientific papers that fail to win the full-throated approval of the referees to whom they have been sent, but which are nevertheless judged to be of sufficient importance to command the interest and attention of readers — sometimes, as in the case of at least two of the referees of the first document in this series, by the sceptics concerned.

The circumstances leading to the publication of the article by Arp, Burbidge, Hoyle, Narlikar and Wickramasinghe nevertheless deserve a little explanation. The authors are distinguished scientists who are known to be iconoclasts. The icon they would destroy is the conventional cosmology known, in shorthand, as the Big Bang model. Their general scepticism is not new. Arp, for

example, while distinguished as an observer, has over decades infuriated colleagues by his insistence that a number of quasars are physically associated with galaxies or other quasars at quite different redshifts. Burbidge, who with his wife, Hoyle and W.H. Fowler was one of the originators of the now standard theory of nucleogenesis, has argued since the discovery of quasars that these objects are not as far away as they seem. Hoyle, who fashioned the modern theory of stellar evolution, was one of the authors of the theory of continuous creation in the 1950s which is echoed in this paper. It is of more than passing interest, even to cosmologists, to know what these people now believe. It will be noted that they have not recanted.

Indeed, their argument may now be more relevant than it would have been a few years ago. The strength of the Big Bang model rests on the simplicity of its concordance with the uniformity of the microwave background radiation and the homogeneity of the Universe on various length-scales. But the mechanism by which galaxies were formed remains almost as much of a problem now as 20 years ago, while the homogeneity of the Universe has been confused by the still-shadowy recognition of large-scale structures. (If the Hubble telescope were functioning as intended, there would be better grounds for hoping that these uncertainties would be quickly resolved.) The difference between Arp *et al.* and their critics is essentially one of judgement. The iconoclasts say that the difficulties listed in their paper give the lie to the Big Bang; their critics hold that this evidence, even if confirmed by further observation, may yet be accommodated within the framework of the Big Bang. It would no doubt be different if Arp *et al.* had now put forward a full account of a more persuasive alternative theory, but that, at this stage, would be a tall order. This journal's position is that such a requirement would be a totally unreasonable precondition of publication, especially when the rest of the community at large will have a vivid interest in what this talented group is about. We are therefore glad to publish this absorbing document, in the spirit of constructive ‘consciousness-raising’.

A word of warning for other authors is appropriate. Arp *et al.* may well complain at the time and trouble they have spent in dealing with particular objections raised by referees. But the outcome is, as always, a more forceful argument — from which the referees still dissent. The most cursory reading of this article will show it to be at once interesting, substantial and couched in the familiar language of science. It is not just another half-baked idea. Others who may regard the new HYPOTHESIS rubric as an opportunity for winning wider circulation for unorthodox ideas should bear all that in mind. They should be warned that they will have to demonstrate the plausibility of what they wish to argue to a small army of sceptics and then convince the editors that the interest of what they have to say outweighs the objections that have been raised. It is not expected that contributions under this rubric will be frequent. □



# Faint hopes for the Faint Object Camera

- Replacement camera an expensive option
- In-orbit repair quick but difficult

**London & Washington**

EUROPEAN space engineers this week launch a study to see what can be done to improve the optical performance of the Faint Object Camera (FOC), built by the European Space Agency (ESA) for the now-crippled Hubble Space Telescope (HST). But Robin Laurance, ESA's space telescope project manager, warns that "the prognosis is not very good" for each of the options being considered.

The FOC was designed to detect extremely faint or distant objects, allowing astronomers to see if distant quasars are embedded in faint galaxies, and to search for planets around nearer stars. Together with the telescope's solar panels, the FOC represents ESA's contribution to the Hubble project. But in common with most of the HST's scientific instruments, the camera's performance has been undermined by the aberration in the telescope's primary mirror.

Laurance says that the ESA study will look at three plans to salvage the FOC, all of them built around a planned shuttle flight to the HST in 1993. Fitting corrective optics to the FOC in orbit is an attractive option, as little observing time would be lost, but it is beset with technical difficulties — corrective lenses may have to be fitted inside the camera, which was never designed to be opened once in space. A safer option would be to bring the camera back to Earth for the necessary corrections, but the refurbished FOC could then not be reinstalled on the HST until the next scheduled shuttle flight in 1996.

The third option is to build an entirely new camera, complete with corrective optics, for launch in 1993. Laurance points out that because of repeated launch delays, the FOC, originally delivered to the National Aeronautics and Space Administration (NASA) in 1983, is now an old instrument. Scientifically, the preferred option would be to start again, but a new instrument would cost about \$100 million, Laurance estimates, beyond the resources of ESA's stretched science budget.

One hope is that a new FOC could be freed from the restricted space science budget and made an ESA 'special project' funded by donations from ESA member states above their mandatory contributions. But this would be an unprecedented move: special project status is usually reserved for commercial ventures, such as the Ariane launch vehicle programme. Another option, suggested by

astronomer Michael Disney, from University College, Cardiff, is for a new FOC to be jointly funded by NASA and ESA.

The FOC is to an extent the victim of its own success. Because its test performance came up to specifications, there were never any plans to build a second generation instrument. In contrast, concern over the detectors in NASA's Wide Field/Planetary Camera (WFPC) means that a second WFPC is already partly built. Fitting corrective optics to this new camera will be relatively inexpensive.

For ESA, the ideal solution would be for NASA to correct the HST's problems at source, in the primary optical system, rather than correcting individual instruments. But the idea of bringing the HST back to Earth was ruled out years ago, according to a NASA spokesman, because of the danger of damage to the telescope and contamination by dust. Similarly, although it would in principle be possible to fit a large corrective lens to the HST as a whole, NASA has ruled against any such plan in favour of replacing scientific instruments one by one on scheduled shuttle flights. The HST's instruments were all designed to be re-

MANNED SPECEFLIGHT

## New blow for Juno

*London*

THE troubled Juno mission, which aimed to send a British astronaut and a series of microgravity experiments to the Soviet Mir space station in 1991, seems doomed now that the Moscow Narodny Bank has stopped looking for sponsors. The original £16 million mission collapsed due to a lack of sponsorship in March, but science director Professor Heinz Wolff had negotiated a rescue package — the Soviets offering a £2 million flight, in return for improved business links with British companies.

Wolff hopes that many of the experiments, into which he has ploughed over £100,000 of his Brunel University laboratory's research contract earnings, will fly on other missions. The bank, meanwhile, is trying to secure a flight for one of the astronauts in training.

Peter Aldhous

placeable by astronauts, and any other repair options would go well beyond the planned serviceability of the telescope.

Meanwhile, astronomers met last week at the Space Telescope Science Institute in Baltimore, Maryland, to begin the difficult task of re-organizing the HST's schedule of observations in the light of what is now known of the telescope's reduced capabilities. The first step will be to assess how much of the currently planned observing programme is salvageable, and to discuss the fate of individual projects with the astronomers who had been granted telescope time for them.

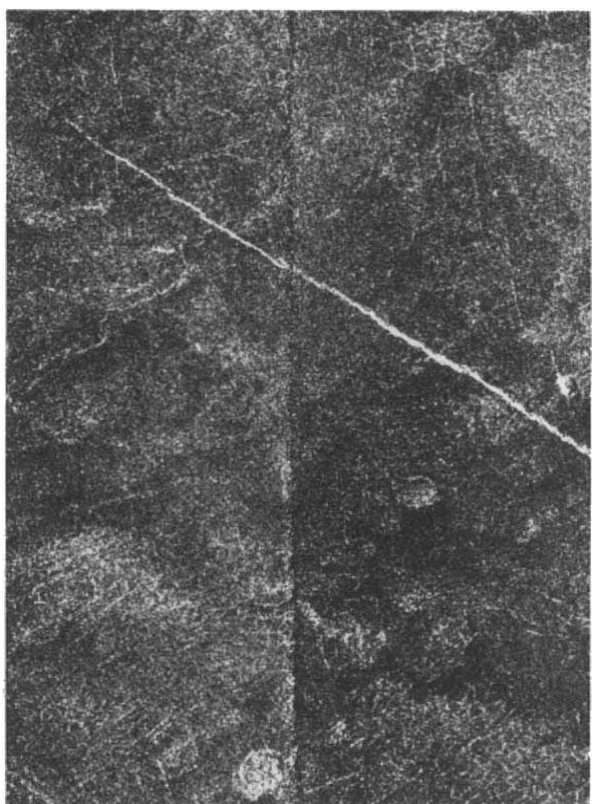
Peter Aldhous & David Lindley

## Computer plays too safe

BETWEEN various mysterious glitches and heart-stopping losses of contact, the Magellan probe last week returned its first radar mapping pictures of Venus. The composite (right) shows one such image, a 24 by 40 mile closeup of a volcanic highland region called Beta Regio. The bright line running across the region is thought to be a fracture or fault zone. Resolution of the picture is about 400 feet, 10 times better than the best previous image of the region from Earth-based radar or earlier probes.

As of late last week, NASA controllers had reestablished contact with Magellan, which has shown a worrisome tendency to "safe" itself and shut off communications for reasons that appear to be connected with the on-board computer. Engineers last week uploaded new software to the probe in hopes of curing the problem.

C.A.





# Researchers take a stand

## London

THE British Association (BA) stepped into the animal rights arena last Friday with its first-ever public declaration of the importance of animal experimentation for the progress of medical research. Prepared over the past few months in response to the increasingly violent attacks of animal rights extremists and included as a late addition to the BA's annual meeting in Swansea, the declaration has the support of a wide array of medical and scientific organizations, including the Medical Research Council (MRC), the Conference of Medical Royal Colleges and the Imperial Cancer Research Fund.

The declaration asserts that animal experiments have made an important contribution to medicine and basic physiology, and that further research "is essential for the conquest of many unsolved medical problems". The scientific community has a duty to explain the aims, methods and benefits of research, it continues, and a responsibility to adhere to animal experimentation legislation and adopt alternatives to animal experiments "as soon as they are proved reliable".

Max Headley, the University of Bristol researcher who narrowly escaped serious injury in an animal-rights car bomb attack earlier this year (see *Nature* **345**, 566; 14 June 1990) hopes the statement will "reassure the public that animal experiments have made a vital contribution [to medicine and veterinary science] and that they must be allowed to do so in future".

Professor Colin Blakemore, a BA vice-president and a distinguished neurobiologist who has been threatened by animal rights campaigners, says he hopes the document will become "a real charter" for researchers. Blakemore regards the growing influence of the animal-rights campaign on public opinion as part of a wider challenge to the role and authority of science. He believes public statements from leading scientific organizations are important to reverse this trend and has long urged that MRC and others should become openly involved in the animal-rights debate (see *Nature* **343**, 3; 1990). Blakemore says that the recent terrorist attacks on animal researchers "have finally brought people off the fence". Ironically, Blakemore says that the declaration, which condemns the violence of animal rights extremists, came on the last day of the BA meeting "because of the fear of reprisals against the local organizers of the conference".

The MRC will in future take a stronger line on the benefits of animal research. A letter from MRC secretary Dai Rees sent to the directors of the council's research units earlier this month says that this year's annual report will contain a state-

ment on animal research, and future news releases on medical developments may include references to the use made of animals in experiments. But the MRC has decided against an aggressive propaganda campaign similar to that adopted recently by the British fur industry.

Moderate animal-welfare organizations give the BA declaration a cautious welcome. Les Ward, director of the Scottish

ANIMAL RIGHTS: USA

## One step forward...

### Washington

AGAINST opposition from animal activists and its own staff, the US Department of Agriculture (USDA) last week released for public comment a set of new regulations for the laboratory care of primates, cats and dogs. The rules, which have been at the centre of a five-year battle both inside and outside the administration, are intended to replace the existing regulations of the Animal Welfare Act. USDA's first attempt to provide new regulations failed when objections were raised by the White House in April (see *Nature* **344**, 804; 26 April 1990).

Perhaps the most controversial change made as a result of criticism from the administration is a clause allowing for "innovative caging" that is not in strict accord with minimum cage sizes. The rules set cage the same as those currently prescribed in the guidelines of the National Institutes of Health (NIH), but make an exemption for "enclosures not precisely meeting the [size] requirements, but that do provide non-human primates with a sufficient volume of space and the opportunity to express species-typical behavior". That, in the language of regulators, is known as a "performance-based" rule and is the principal sign of the defeat of USDA, which had insisted that cage parameters should be precisely specified. The winner appears to be the White House, which argued that the rules proposed earlier would have bankrupted the research community. USDA estimated that those rules, with strict minimum cage sizes, would have cost research facilities \$1,750 million. Costs to facilities are reduced to \$552 million in the new version.

But USDA officials say that the changes may also make the rules unenforceable. If every laboratory describes its caging as 'innovative', inspectors will no longer be able simply to take out a tape measure to determine infractions, they argue. Instead, rulings on the adequacy of cages will turn on the opinion of other scientists. "You're pitting one expert against another", says Richard Crawford, director of the USDA animal care staff.

group Advocates for Animals says the statement is "well thought out", and hopes that scientists will pursue the alternatives to animal experiments mentioned in the statement "with vigour". But Michael Balls, from the Fund for the Replacement of Animals in Medical Experiments, says that this has not often been the case in the past. He says the signatories must now be "as active in supporting the development of alternatives to animal experiments as they are in defending animal research". **Peter Aldhous**

"We're going to have to go to experts wherever they are, and that usually means industry. If industry just circles the wagons, this isn't going to work." Animal welfare activists are no happier. "By including the innovative caging specification they've made every 'shall' a 'should'", says William Cotreau of the Animal Welfare Institute.

But Frankie Trull, president of the National Association for Biomedical Research, says that "innovative" cages are more likely to be the exception than the rule. Even when inspectors are forced to make judgements, the condition of the animals should be the most important consideration, she says.

James Wyngaarden, former NIH director, who until recently handled biomedical issues for the White House Office of Science and Technology Policy, says the new regulations are "a step in the right direction" towards the performance-based rules he had advocated in several critical memoranda to USDA.

The regulations also say that laboratories must "develop, document, and follow a plan" for the psychological well-being of primates. Although the proposed regulations lay out certain objectives, details of the plan are left to each laboratory.

Cotreau, noting that USDA inspectors were able to inspect only a quarter of all dog dealers in 1988, says the additional responsibility of determining what the well-being plan is, whether it is sufficient and whether it is being followed could prove impossible for the overtaxed inspectors. USDA's Crawford says that maintaining a reasonable inspections schedule would probably require more inspectors. But given budget constraints, that could prove difficult, he says.

Unlike the regulations for rabbits, hamsters and guinea-pigs issued last month, the primate and dog regulations have no "grandfather clause". Existing caging is thus included along with new enclosures. USDA has set a deadline of 1 October for comments before preposing a final rule. **Christopher Anderson**

# A darker shade of green

## Tokyo

RESEARCH on the global environment and space feature prominently in budget requests for 1991 to be submitted on 31 August by Japan's Science and Technology Agency (STA) and the Ministry of International Trade and Industry (MITI). The requests, for overall increases of 6 per cent (STA) and 4 per cent (MITI), are thought likely to be reduced only slightly in negotiations with the Ministry of Finance before going for Diet approval early next year. The really critical negotiations are already over, taking place each year just before the request goes to the Ministry of Finance.

Both MITI and STA have plumped for large increases for research on the global environment. MITI requests ¥7,153 million (\$48 million), more than 17 per cent up on this year. A large part of the budget will go towards development of "environment friendly" technology, such as CFC substitutes, biodegradable plastics, and technology to absorb and utilize carbon dioxide. These projects will at first be carried out in MITI laboratories and private industry. But most of them will be transferred to a new institute, the Research Institute of Innovative Technology for the Earth, which is expected to open in 1992 in the new Kansai science city between Kyoto, Nara and Osaka.

Kyoto prefectural government has provided free land for the institute and will also level the site, which is in a steep valley next to MITI's new Ion Engineering Center (see *Nature* 346, 210; 1990), free of charge. And Ikuro Tomita, director for global environmental technology in MITI's Agency of Industrial Science and Technology (AIST) expects the local government and private industry to contribute an additional ¥8,000 million or so towards construction and operation of the institute.

MITI continues to increase its budgets for development of energy-saving technology and alternative energy sources under the 'Moonlight' and 'Sunshine' projects. But the ministry has decided to decrease dramatically the budget for coal liquefaction and gasification from ¥24,901 million this fiscal year to ¥16,296 million because the technology will not help reduce carbon dioxide emissions. As

a result, the overall budget for Sunshine drops by nearly 24 per cent.

STA makes an ambitious request of ¥10,532 million (\$70 million), over double the budget for this year, for its 'green planet project' which aims at better predictions of global warming. More than half of the budget (¥5,449 million) goes towards development of the ozone-sensing Advanced Earth Observing Satellite (ADEOS), due to be launched in 1995.

Also included in the request is ¥354 million for a supercomputer to model climate change. According to an STA official, the agency will select a Cray (the Japanese government has been heavily criticized by the United States for its failure to buy US-made supercomputers).

ADEOS is provided for by one part of an 11 per cent increase for space projects. Most of the rest of the extra money goes to Japan's much-increased contribution to the US space station. The agency requests ¥18,000 million, nearly double the budget

### JAPANESE SCIENCE BUDGET REQUESTS

| Ministry of International Trade and Industry (MITI) |              |                      |
|-----------------------------------------------------|--------------|----------------------|
|                                                     | ¥000 million | (% change from 1990) |
| Total R & D budget                                  | 260.1        | + 4.0                |
| Japan Key Technology Centre                         | 32.2         | +23.9                |
| Basic technologies for future industries            | 8.0          | + 6.7                |
| Large-scale industrial projects                     | 14.6         | + 3.5                |
| Sunshine project                                    | 30.1         | -23.8                |
| Moonlight project                                   | 12.3         | + 5.7                |
| Unmanned space platform                             | 5.7          | + 7.5                |
| Fifth-generation computer                           | 7.3          | + 4.3                |
| Sixth-generation computer                           | 0.1          | —                    |
| Global environment                                  | 7.2          | +17.4                |
| IMS                                                 | 0.1          | +36.4                |
| Science and Technology Agency (STA)                 |              |                      |
| Total R & D budget                                  | 524.6        | + 6.0                |
| Special promotion funds                             | 10.6         | + 3.9                |
| Space                                               | 132.3        | +10.8                |
| Nuclear energy                                      | 305.3        | + 3.0                |
| SOR (SPRING-8)                                      | 4.9          | +75.0                |
| Ocean research                                      | 10.7         | + 8.1                |
| ERATO                                               | 5.8          | +12.8                |
| Human Frontier Science Programme                    | 4.1*         | +28.2                |
| Green Planet Project                                | 10.5         | +123.7               |
| Sakigake Research 21                                | 0.6          | —                    |

\* Includes ¥1,665 million from MITI budget

for this year. The project will continue to be a large drain on STA's budget over the coming decade.

MITI also continues to pour quite a large budget (¥5,700 million) into development of an unmanned space platform in collaboration with STA and the Ministry of Education, Science and Culture (MESC). The platform is scheduled for launch in early 1994 (provided there are no more hitches with the H-II rocket).

Ocean research also benefits from concern about the Earth with an increase of over eight per cent under STA's budget, although the increase is considerably less than STA's Ocean Development Division hoped for. Most of the extra money will go to research on the deep sea and to investigation of the role of the ocean in climate

change.

STA's budget for the world's largest synchrotron, SPRING-8, continues to grow and the project will consume about ¥100,000 million by the time of its completion in 1998. The prototype fast-breeder reactor Monju also gulps down a huge budget (¥47,100 million) as its expected completion in 1992 approaches. And the ill-fated nuclear ship *Mutsu* which is due to be scrapped next year after spending only a few months at sea (and 22 years in port), gets ¥3,855 million. Most of the money will probably go to decommissioning costs.

MITI and STA continue to expand their budgets for the Human Frontier Science Program which supports international research on the brain and molecular biology. With the increased budget, the Frontiers foundation in Strasbourg will be able to award at least 30 three-year grants every year worth on average about ¥38 million (\$250,000) a year, even if contributions from the other summit nations (Canada, Britain, France, United States, Italy and West Germany) and the European Communities (EC) remain small.

A surprising development is a decision by STA to increase its budget for research on the human genome by over 60 per cent to ¥1,000 million (\$7 million). Earlier this year, STA officials said they did not intend to expand the agency's genome budget. But even with the increase, Japan's contribution to genome research remains tiny compared with that of the United States. MITI officials are considering ways of establishing their own genome project but it is likely to be at least a year before there is any chance that it can be launched.

STA's budget for ERATO continues to grow, with four new projects scheduled to start next year. And the agency will launch a new programme next year, Sakigake Research 21, modelled along the lines of ERATO but intended to support individual researchers rather than teams. The aim is to provide three-year contracts each worth on average about ¥20 million (\$135,000) a year to about 45 researchers from universities, government laboratories, and industry, probably in bioscience, materials science and electronics.

MITI requests ¥110 million to prepare a follow-up to the fifth generation computer project which ends next fiscal year. The new project will probably involve the development of neurocomputers and optical computers. MITI's budget also includes a similar amount (¥150 million) for the Intelligent Manufacturing System (IMS) project which the ministry hopes to launch in collaboration with the United States, Europe, Australia, and possibly Canada. Officials of the US Department of Commerce, the EC, and Japan will discuss IMS at a meeting in Tokyo next month.

David Swinbanks



# Putting faith in theory

## Seattle, Washington

ERNEST Henley, director of the University of Washington's new \$7.5 million centre for nuclear theory, says that theoretical physicists are in short supply. Never mind that young theorists, unable to find a job in their profession, have been leaving the field in droves.

Physics needs more theorists, not fewer, Henley says. While this may seem counterintuitive (and, indeed, many physicists disagree), it is also the conclusion of a 1988 report by the Department of Energy's Nuclear Science Advisory Committee, whose chief recommendation — the establishment of the centre that Henley heads — has now been implemented.

The problem, Henley says, is funding for theoretical physics, not the demand for theoreticians. And the situation is nowhere worse than in the area of nuclear physics. The Department of Energy report noted that theorists represent only 18 per cent of all nuclear physicists, as compared to 47 per cent of particle physicists. Only in optics, traditionally a relatively applied field, is the proportion of theoretical physicists lower. The panel recommended that the Department of Energy add \$38 million over five years to its budget for nuclear theory, including some \$7 million for a centre, to revive the field.

The Department of Energy, which had just committed itself to two major new nuclear physics facilities (RHIC, the Relativistic Heavy Ion Collider at Brookhaven National Laboratory and CEBAF, the Continuous Electron Beam Accelerator Facility) endorsed the plan. After a national competition, the University of Washington was selected as the site of the new theory centre. Funding was included in the 1990 budget, and the centre opened its doors in March.

Top priorities for the new centre are to spur collaborations between nuclear theorists and physicists in other fields, and to encourage graduate students and postdoctoral researchers to enter and stay in theoretical physics. The methods to be used, at least for the centre's first series of programmes this summer, appear to be basically similar to those at existing theory centres such as the National Science Foundation's Institute for Theoretical Physics at the University of California (UC), Santa Barbara. A group of between 10 and 20 scientists are invited to spend two or three months at the centre, to participate in daily hour-long seminars and informal brainstorming sessions. The object, of course, is to focus on problems that lend themselves to collaboration, and like any good dating service, the centre will be judged by the number of marriages

it arranges. Several joint papers per session is an early goal.

Despite its interdisciplinary ambitions, the University of Washington centre is unique in its foundations on nuclear theory alone. Although the UC Santa Barbara centre includes nuclear theory in its charter, actual research there tends to be concentrated in condensed-matter and particle physics. The lack of research in nuclear theory has been largely due to a funding shift in the 1980s away from both nuclear physics and theory in general, says Henley. But the pendulum may be swinging back again. Many new problems in particle physics and astrophysics hinge on classic nuclear theory. Quark and gluon propagation in very dense matter, for example, turns out to be a many-body problem, raw meat for a nuclear theorist. "We're going to show the community the connections between nuclear theory and other disciplines", he says.

But a rediscovery of nuclear theory's importance will pay off only if there are good theorists to call on. Attracting young

talent to the field is where the centre is most likely to make a difference. Half of the invited participants in each session are graduates or young postdoctoral researchers. And unsolicited applications from prospective theorists are likely to be viewed with favour. Five full-time staff will be young scientists on five-year appointments. That should tide them by for the first five years after their degree — the years in which they are most likely to leave the field in frustration after trying and failing to get a job — and bring them to the attention of potential employers.

As massive projects such as the Superconducting Super Collider (SSC) draw an increasing number of physicists into multi-year experiments with hundreds of participants, theorists will become even more important in planning new projects, Henley predicts. "New experiments for particle physicists will come maybe once every six years. They're going to need good theorists to suggest the next experiment and to interpret the last one." With a decade to go before the SSC is up and running, the Department of Energy's new theory centre may be in the right place at the right time. **Christopher Anderson**

## SWISS SCIENCE

### New president to promote ETH

#### Munich

THE Swiss Federal Institute of Technology, better known by its acronym ETH (for Eidgenössische Technische Hochschule) has a new president in microbiologist Jakob Nüesch. Nüesch, 57, a graduate of the institute, returns to academic life after



Busy man: Nüesch is looking for new professors at a rate of two a month over the next four years.

28 years in industry, most recently as director of pharmaceutical research at Ciba-Geigy in Basel.

Nüesch's appointment follows the surprise early resignation in March of mathematician Hans Bühlmann. In three years served of a scheduled five-year term, Bühlmann made several changes in the structure of ETH, introducing a department system and expanding the role of

research by appointing a vice-president for research.

Emotions still run high regarding the changes brought in by Bühlmann. According to one professor, Bühlmann and his vice-president for finance, computer scientist Carl Zehnder, closed themselves off from the faculty and it seemed that "they were basically interested in seeing if they could run ETH with a computer". This professor feared that the three-year tenure of Bühlmann would leave lasting scars on ETH because of the loss of several key administrators who "cared about research". But the more sympathetic faculty members, such as biochemist Kaspar Winterhalter, say that Bühlmann and Zehnder became the undeserving scapegoats for the faculty's frustrations at bureaucratic delays and a "grotesque" shortage of space.

In addition to tackling the space problem, which he admits is urgent, Nüesch faces several daunting tasks. He will have to recruit a hundred professors in the next four years, who will shape research at ETH well into the next century. He would also like to begin to increase the number of foreign students (currently 14 per cent) and women (less than 30 per cent) at ETH. Nüesch would especially like to attract more students from European Communities member states, which he hopes will strengthen the bonds between Switzerland and the European Communities.

Steven Dickman



## NUCLEAR BOMB TESTS

# Still dirty after all these years

## Sydney

DESPITE earlier clean-up attempts, plutonium still contaminates the Maralinga ('Fields of Thunder') test site, northwest of Adelaide, which was used by the British during the 1950s and early 1960s for above-ground nuclear bomb tests. According to a report that will be tabled in the Australian parliament later this session, complete removal of the remaining radioactivity could cost as much as A\$600 million, a sum towards which the British government may be asked to contribute.

Following a recommendation in 1985 by a Royal Commission that Maralinga should be cleaned up and returned to its original inhabitants, the Pitjinjara aborigines, a Technical Assessment Group (TAG) was set up in February 1986 by the federal Department of Minerals and Energy to assess the level of contamination and estimate the cost of decontamination. The TAG was a joint Anglo-Australian venture, with independent representation from the United States.

The TAG identified an area of 95 square kilometres over which radioactivity levels due to plutonium were five times greater than the internationally recognized 1 millisievert (mSv) safety standard. According to the executive summary of the TAG report, which was leaked to an Australian television network, full decontamination of the affected area would require collection and reburial of soil, and would cost A\$600 million. Less expensive but less thorough options described by the TAG include ploughing to dilute the surface soil, or overlaying the present surface and increasing security around the site. The present system of security fences costs A\$7 million a year to maintain.

According to Rob Robotham, director of safety and health management at the University of Melbourne and author of a book on Maralinga, the sums of money being considered for clean-up would be better spent elsewhere. He says that if there were 1,000 people living on the site the added health risk, chiefly from inhalation of dust and ingestion of plants and animals contaminated with plutonium, would amount to one case of cancer, probably of the lung, in 20 years.

Complicating the debate over what should be done is the question of who should pay. In 1967, the Australian government released the British from further liability after they had completed 'Operation Brumby', in which contaminated areas in the vicinity of large explosions were cleaned up. But according to Robotham, the British neglected areas affected by what were deemed "minor tests". He also says that breaking up the soil, which the British thought would solve contamination problems, in fact spread it around.

## IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Explosion at Maralinga in 1957. The device was suspended 1,500 feet in the air. (AP)

Jim McClelland, the judge who was in charge of the Royal Commission that recommended clean-up, says the British have a responsibility to return Maralinga to a habitable state. He says that when Harold Holt [then prime minister] released the British government from further responsibility, he "didn't have all the information about the extent of the plutonium dispersal. There is now a legal and an overwhelming moral obligation on the part of the British to clean the place up." The British High Commissioner, however, responded with the statement that "Mr McClelland can say what he likes...it's a matter for the Australian government".

Tania Ewing

## ENVIRONMENTAL RESEARCH

## Japan to help China to clean up

### Tokyo

JAPAN will help China to fight its chronic pollution problems by providing Overseas Development Aid (ODA) to build an environment research institute in Beijing, said an official at Japan's Environment Agency last week. The idea of an ODA-funded institute was first proposed by China in 1988, but Japan suspended discussions after the military crackdown in Beijing in June last year. At the recent summit of Western nations in Houston, Texas, however, Japan won an understanding to resume aid to China, opening the way for renewed discussions on the institute.

In a meeting in Beijing last month, Japan agreed to provide ¥10,000 million (\$67 million) for the institute, subject to approval by Japan's Ministry of Finance, and to transfer antipollution technology from Japan. Construction of the institute will begin in the second half of next year and is expected to be completed in 1994, according to an environment agency official.

China has severe pollution problems. In particular, emission of sulphur dioxide from coal-burning is believed to be causing acid rain to fall as far afield as the west coast of Japan (see *Nature* 340, 671; 1989). Japan will provide China with technology and expertise to reduce such emissions, but details of the activities of the institute will not be decided until another meeting between Japanese and Chinese officials later this year, according to an official of the grant-aid division of the Ministry of Foreign Affairs. David Swinbanks

## RABIES VACCINE

## Better late than never for start of tests

### Washington

THE first field tests of a genetically engineered rabies vaccine are at long last under way in the United States—some four years after field tests were conducted in Canada and months after a full-scale campaign to eradicate rabies from the fox population began in France. Cumbersome regulatory processes in the United States are to blame for the delay, according to Warren B. Cheston, associate director of the Wistar Institute which helped develop the vaccine.

Rabies is now at epidemic proportions in the raccoon population in the mid-Atlantic states and there are fears it could spread to domestic animals, including pet cats and dogs. The vaccine is a recombinant vaccinia-virus vaccine and is the product of collaborative research between the Wistar Institute of Philadelphia and Transgene of Strasbourg, France.

Trials are being carried out on the uninhabited Parramore Island off the Virgi-

nia coast. The vaccine is placed in a bait containing fish oils that are abhorrent to man and other animals but attractive to raccoons. Blood tests will be carried out for up to a year on raccoons to determine the level of protection against rabies.

Federal approval for the field test, which was originally to have taken place in South Carolina, was granted by the US Department of Agriculture (USDA) in March 1989, but Wistar failed to obtain the approval of the South Carolina public health authorities. It then moved the project to Parramore Island, and received approval in July 1989. A further year was consumed by negotiations with the Nature Conservancy which owns the island.

Jane Rissler, spokeswoman for the National Wildlife Federation, said that it was only when the Nature Conservancy became involved that "the protocol finally came up to the science and safety standards that the USDA should have imposed".

Diane Gershon

# LA to Vegas in 75 minutes

## San Francisco & Munich

AFTER years of development in several countries, the world's first long-distance magnetically levitated train (maglev) appears to be on the way to reality by 1997 in the western United States.

Although short prototype maglev lines have been in operation in West Germany and Japan for several years, a 427-km passenger line between the Los Angeles area and Las Vegas, Nevada, using German train technology, will be the first truly commercial venture of its type.

The go-ahead came two weeks ago when the California-Nevada Super Speed Train Commission awarded a franchise for construction of the maglev line to a consortium led by Bechtel International, a large engineering concern, of San Francisco. Transrapid International of West Germany will provide maglev technology in what is its first commercial contract. Approval of the franchise is being termed "tentative", but both the consortium and the commission are confident that final approval will be granted in the next few months. Construction should then begin in 1993.

Bechtel expects a total cost of \$5,000 million for the line, which will run to Las Vegas from Anaheim, in the Los Angeles metropolitan area. The line will be financed solely through private investors. Other consortium members include the C. Itoh Company of Japan, which is expected to act as a lead representative for Japanese investors, and Amtrak, the US passenger rail operator, which will provide operations, maintenance and marketing services.

In Transrapid's maglev technology, a U-shaped track carries a varying electromagnetic field that acts on conventional magnets underneath and along the sides of the train, supporting it and drawing it along. In contrast, the Japanese maglev system uses superconducting magnets, but is not yet ready for commercial use because, according to Bechtel vice-president Erv Koenig, a method to shield passengers from the very strong magnetic field has not yet been developed.

The Bechtel and Transrapid teams are also ready to proceed by November on construction of a short maglev line in Florida, which will run from Orlando to Disney World. Despite the US selection of its system, Transrapid still sees the Japanese alternative as a "very serious competitor", according to Thomas Naujoks of Thyssen Henschel, leader of the Transrapid consortium.

Transrapid has had less success at home: it has not yet found funding to build a 55-km test track which was approved last December to run between Düsseldorf and the Cologne-Bonn airport.

Journey time between Las Vegas and

Anaheim is estimated at only 75 minutes, the trains maintaining speeds of over 500 km per hour for most of the distance. Its developers claim the train will make only about as much sound as an automobile as it passes, and that the electric powered maglev will replace 620,000 automobile-kilometres each year, equivalent to a reduction of tons of smog-causing nitrogen oxide emissions.

The US line will be based on a prototype train currently operated by Transrapid on a 31.5-km test track in Emsland, West Germany. That train logged 100,000 km this month, and data generated from test runs are now being examined by German and US railway regulatory officials. Koenig says he expects the certification required for construction to be made by 1 October.

This latest date and details of Bechtel's

## RESEARCH FACILITIES

# Economics lesson for MIT

## Washington

SCIENTISTS at the Massachusetts Institute of Technology (MIT) may have lost a sure-bet \$118-million magnet laboratory last week, but they did get a valuable lesson in the politics of science funding. To wit: money talks, and \$58 million talks very loudly.

Painful proof arrived on 17 August, when the National Science Foundation (NSF), with the blessing of its National Science Board, chose Florida State University as the future site of the National High Magnetic Field Laboratory (NHMFL), a user facility that will produce the world's most powerful steady magnetic fields by the time of its completion in 1996. Two independent review panels had picked MIT (whose Francis Bitter Laboratory is the leading US steady-state magnetic field facility) as the better of the two sites, but when the state of Florida sweetened its pot with a \$58-million matching offer, NSF found it hard to say no.

"Our responsibility is to get the biggest bang for the buck", explains David Sanchez, NSF's assistant director for mathematical and physical science. MIT was able to offer industrial contributions of only about \$10 million to the project. Sanchez notes that, although the scientific review panels favoured MIT, they also concluded that "the US would be well served by either proposal". Although Florida State has no existing magnet facility, it has arranged a partnership with the Los Alamos National Laboratory and the French Grenoble Magnet Laboratory, both of which have world-class magnet programmes. The university has also promised to create 54 new faculty and staff

financial feasibility studies will be presented to the commission at their next meeting. Both sides indicate they expect negotiations of the conditions for final approval to go smoothly and to be completed within two to four months.

Bechtel officials have presented their maglev plan to several environmental groups already, and have encountered no serious opposition. Construction plans call for the train to run beside an existing major highway, so no new environmental disruption will occur. The tracks are to be elevated and supported by pillars so that the natural environment can reestablish itself between supports.

Commission chairman Arnie Adamsen expects the largest segment of the train's ridership to consist of tourists, with "recreational" riders and commuters providing additional passengers. The round-trip fare is likely to be \$110 at present prices, 30 per cent less than airline fares.

**Elizabeth Schaefer & Steven Dickman**

positions for the laboratory. Sanchez expects that in five years Florida State will be at the same point the MIT laboratory would have been and will have surpassed the predicted MIT position within ten years.

NHMFL will be the United States' top magnet laboratory, and is to feature a direct-current magnet generating a field of 45 tesla, almost 10 tesla higher than the current world record (held by MIT). Intense magnetic fields can be used to explore the structure of complex biological molecules, and to create new materials.

Paul Gray, president of MIT, released a statement last week announcing that the university will ask NSF to reconsider the decision, and is filing an appeal based on scientific and procedural arguments. Although Sanchez says NSF will consider any request, he points out that all the National Science Board did was to approve an internal NSF decision to pick Florida State. There is no formal appeal process for the decision. Sanchez says that NSF will continue funding the Francis Bitter Laboratory for at least one more year at a level of \$6 million.

The NSF decision is seen as a sign of a growing trend towards pragmatism in an increasingly tight fiscal climate. "First Texas buys the SSC [Superconducting Super Collider] and now Florida buys the magnet facility. Is this going to be the rule in the future?" ask Robert Park, director of the American Physical Society's Washington office. Park is concerned that universities that are not part of a consortium or do not have substantial state support may find it increasingly difficult to win large projects. **Christopher Anderson**



## GENETIC ENGINEERING

# Tryptophan under suspicion

## Washington

DEBATE over the safety of recombinant-DNA technology is heating up again following the publication of a paper in the *New England Journal of Medicine* that describes a possible link between an outbreak of eosinophilia-myalgia syndrome (EMS), a sometimes fatal blood disorder, and the consumption of a genetically engineered amino acid. Jeremy Rifkin, president of the Foundation on Economic Trends, a Washington-based group that has long opposed the use of recombinant-DNA technology, has quickly seized upon the new data and petitioned the US Food and Drug Administration (FDA) to stop the licensing of all new genetically engineered products until a full public inquiry has been carried out.

The product implicated in the EMS outbreak, which affected more than 1,500 Americans and led to 27 deaths, was L-tryptophan, a dietary supplement commonly taken to alleviate insomnia, depression and premenstrual tension. Although six companies manufacture L-tryptophan and import it into the United States, Michael Osterholm of the Minnesota Department of Health and his co-workers report in the *New England Journal of Medicine* (323, 9 August 1990) that the outbreak of EMS can be linked to batches of L-tryptophan manufactured in early 1989 by Showa Denko, Japan's fourth-largest chemical company.

Batches of L-tryptophan produced in these months are thus assumed to have contained some chemical impurity that contributes to the development of EMS. Although the researchers do not know what the impurity is, they point out that two changes in procedure took place at Showa Denko around that time. First, the company changed its purification process, reducing by half the amount of carbon used to filter out contaminants. Second, the company introduced a new genetically engineered strain of *Bacillus amyloliquefaciens* — Strain V — in the fermentation process.

Osterholm and his co-workers are not able to say which of these two factors might have been responsible for the appearance of the postulated impurity. The FDA backs them up, saying that it is impossible to conclude that the use of a "bioengineered organism" was the "key causative factor behind the epidemic".

Showa Denko voluntarily removed its product from the US market in November 1989 when preliminary medical data were released. A company spokesman confirmed that more than 100 law suits have so far been filed against Showa Denko in the United States.

Rifkin has formally requested that the FDA take a "middle course" of action,

which would leave approved drugs and products produced by recombinant-DNA technology on the market but which would suspend the review of all new products until certain measures are completed. The petition calls for a risk-assessment study by the FDA of the dangers of recombinant-DNA technology, full public disclosure of its findings so far in the inquiry into the deaths and illnesses associated with L-tryptophan and a re-evaluation of FDA's policy regarding

## EASTERN EUROPEAN ENVIRONMENT

# Cleaning Kola with kroner

## Munich

NORWAY has set in motion a plan of unprecedented size and cost to help ease environmental pollution coming from a remote corner of Eastern Europe. Financial aid and loan guarantees are to go from Norway and other Scandinavian countries to the Soviet Union so that it can replace antiquated nickel smelters at Nikel and Zapolyarnyi on the Kola Peninsula, just a few kilometres from the Norwegian border.

The arrangement is of exactly the type

the regulation of biotechnology products.

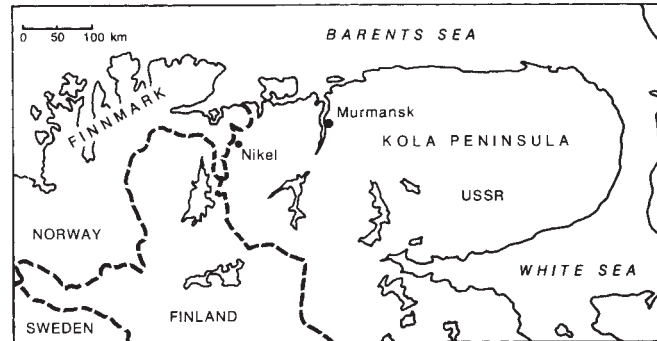
Industrial Biotechnology Association president Richard Godown warned against overreacting, and said that it is quite possible that the problem is the result of "a purification foul-up". Any move by the FDA to withhold valuable therapeutic medicines from the public, as suggested by Rifkin, would be "extremely imprudent", he says.

Rifkin has threatened legal action if the FDA does not respond with affirmative action within 20 days, but believes the FDA will meet enough of the demands to avoid litigation.

**Dlane Gershon**

fouled the air and soil, turning several square miles of the sparsely populated arctic region into a virtual desert. When Norway was finally permitted to send a team to monitor the environment on the Soviet side of the border last year, it found conditions to be as bad as anywhere in Eastern Europe, with the critical difference that the fragile soil had a much lower capacity to buffer acid depositions. "All our observers were shocked by what they saw", says Per Antonsen of the Norwegian Environment Ministry. "Pollution far

exceeds the critical levels that the environment can absorb." Norwegian officials have also noticed "anomalies" in the incidence of cancer and severe allergic reactions to nickel in Finnmark, but the population of the area is too small for firm conclusions, he says. A study of Soviet



advocated by some environment analysts who study transboundary pollution, including experts at the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria. Although their advice has so far been ignored, they have said for years that Western Europe would be better off investing its anti-pollution funds directly in Eastern Europe, rather than in trying to ameliorate the effects of pollution in the West.

Acid emissions from the smelters have caused an ecological disaster in Norway's northeasternmost county of Finnmark. The smelters were built in the 1930s and have operated without pollution controls ever since, dumping nearly 600,000 tonnes of sulphur dioxide a year onto the Soviet Kola Peninsula and neighbouring Norway and Finland.

The sulphur emissions, along with 11,000 tonnes a year of heavy metals, have

workers at the plants would reveal a significant effect of pollution on health, Antonsen believes.

The Norwegian government expects to complete negotiations for the clean-up in October, says Antonsen. Norway has already promised to pay Nkr 300 million (about \$50 million), and is said to have arranged bank credit of a further Nkr 700 million with the Nordic Investment Bank in Helsinki. The Finnish government and perhaps Sweden and Denmark are also expected to participate in the financing.

The overall clean-up will cost much more, perhaps as much as Nkr 4,000 million, says Antonsen, of which the Soviets will pay as much as 60 per cent. The Norwegian people stand firmly behind the government in paying for the clean-up, says Antonsen, because they know that reducing Soviet sulphur emissions is the most efficient way to eliminate pollution in Norway.

**Steven Dickman**



## Problems of Rome University

SIR—I am one of the associate professors of Rome University who, according to Professor Paolo Amati *et al.* (*Nature* **345**, 658; 1990) publish about 0.1 papers a year in “international refereed journals”. I do not know where my colleagues have collected these data, but they seem to refer to the year 1989.

As far as I know, there is no law that compels a prospective professor to learn a foreign language so well as to publish papers in that language. In addition, most Italian scientific journals are reviewed by the *Index Medicus*, *Current Contents*, *Excerpta Medica* and so on. For example, the *Rivista Italiana di Pediatria* (*Italian Journal of Pediatrics*) is reviewed by *Current Contents/Clinical Practice* and *ISI/Biomed*; the *European Review for Medical Pharmacological Sciences* (which accepts papers also in French, German, English and Spanish) has been included in both the *Index Medicus* and *Excerpta Medica* reviews. In addition, most Italian journals require English abstracts and keywords, and publish contents in English.

My colleagues are all professors of science, whereas many Rome University professors belong to clinical/surgical dis-

ciplines. In addition to lectures, examinations and scientific research, a clinical professor must devote time to patients, and a surgeon also has to perform operations. They both have heavy teaching loads in postgraduate medical schools. Associate professors in all Italian universities receive the same salary.

I have no comment on the decision not to appoint younger colleagues. But in every free assembly, from parliament to faculty meetings, everybody can express an opinion and, eventually, the majority approves or rejects. But if a proposal is democratically rejected, it is not fair to complain.

Perhaps I am not a black sheep in the eyes of Amati *et al.* Although I have been assistant and then associate professor at Rome University since 1962, I have not even a stool in the pediatric department where I have worked since 1964. What has frustrated me has been the refusal of either the university rector or the faculty dean to help.

ARNALDO CANTANI

*Istituto di Clinica Pediatrica,  
University of Rome “La Sapienza”,  
00161 Rome, Italy*

## Duesberg replies

SIR—Weiss and Jaffe<sup>1</sup> say that I am a “flat-earth bogged down in molecular minutiae and miasmatic theories of disease” for not accepting their hypothesis that AIDS is infectious. If they are right, I would indeed be “perilous”<sup>1</sup> for questioning AIDS-prevention through safe sex.

As we agree that HIV (human immunodeficiency virus) fails Koch’s postulates<sup>1,2</sup>, Weiss and Jaffe could prove me perilous by showing that: (1) safe sex has stopped AIDS in any country; (2) among matched groups of homosexuals, haemophiliacs or intravenous drug users, only those infected with HIV get AIDS; and (3) the mortality of American haemophiliacs has increased since 16,000 of them (80 per cent) became infected around 1980. Where is that proof?

And where is Weiss and Jaffe’s “strong evidence against lifestyle explanations”, if 97 per cent of all people with AIDS in the United States are from health-risk groups: 30 per cent intravenous drug users, 60 per cent homosexuals and heterosexuals, who frequently use oral aphrodisiacs and psychoactive drugs (nitrite inhalants, cocaine); and 7 per cent haemophiliacs, recipients of transfusions and babies of drug-addicted mothers<sup>3,4</sup>?

And why does a sexually transmitted disease remain almost exclusively (more than 90 per cent) male — now for nine years, in a free country like the United States<sup>4</sup>? And why does HIV-neutralizing

antibody (the AIDS test) indicate a virus that has yet to become pathogenic? And why is the cytotoxic DNA chain terminator AZT prescribed to sick and healthy persons who carry a dormant provirus in only 1 of 500 lymphocytes<sup>2,3</sup> — a “perilous” attack on “molecular minutiae”?

As a virologist, I understand Weiss and Jaffe’s fascination with infectious agents. But it is romantic, not scientific, to abandon proven rules such as Koch’s postulates without providing new ones to accommodate HIV as the cause of AIDS. Thousands of lives have been lost in the past because medical scientists, inspired by successes of the germ theory, have misdiagnosed as infectious diseases vitamin deficiencies, such as pellagra in the United States<sup>5</sup> or scurvy in the United Kingdom<sup>5</sup>, and recently a drug-induced neuropathy in Japan<sup>6</sup>. Because the virus-AIDS hypothesis has yet to save a single life, ridiculing a testable alternative<sup>3</sup> might be “perilous”.

PETER H. DUESBERG

*Department of Molecular and  
Cell Biology,  
University of California,  
Berkeley, California 94720, USA*

1. Weiss, R.A. & Jaffe, H.W. *Nature* **345**, 659–660 (1990).
2. Duesberg, P.H. *Proc. natn. Acad. Sci.* **86**, 775–764 (1989).
3. Duesberg, P.H. *Res. Immun.* **141**, 5–11 (1990).
4. US Dept of Health and Human Services *HIV/AIDS Surveillance* (Centers for Disease Control, January 1990).
5. Bailey, H. *The Vitamin Pioneers* (Rodale, Emmaus, 1985).
6. Kono, R. & Shigematsu, J. *Jap. J. Med. Sci. Biol. Suppl.* **28**, 1–21 (1975).

## Role of religion

SIR—I would be surprised if there should be any field of enquiry not open to the pages of *Nature*. Religion has important sociological functions and is therefore susceptible to scientific analysis. While Christopher J. Lote may consider religion to be destructive (*Nature* **346**, 10; 1990), there is evidence that religion has a role in maintaining social structures and establishing patterns of behaviour with concepts of fairness, justice and altruism. To dismiss any discussion of religion would seem premature when the data have rarely been objectively assessed.

In a straw poll conducted in this hospital of medical staff, 13 out of 20 had a religious belief, a prevalence significantly greater than found by Lote. Small studies are probably unhelpful in establishing whether there is any relationship between science and atheism.

Science and faith in God need not be contradictory, as Lote proposes. Religions of the Judaeo-Christian tradition would appear to suggest that observation can substantiate their claims (Psalm 19:1–4; Romans 1:20). Is the existence of God a testable hypothesis? Christians would claim that performing the appropriate experiment (Romans 10:9–13) will result in revolutionary and therefore observable changes (1 Corinthians 5:17).

GRAHAM BOTHAMLEY

*Department of Thoracic Medicine,  
London Chest Hospital,  
Bonner Road,  
London E2 9JX, UK*

SIR—I hope you will run an opinion poll such as that suggested by Christopher Lote. In the meantime, can your regular reader and occasional correspondent God, through one of His/Her agents, put my mind at rest on the following matter? In the United Kingdom, we have what is called the national anthem, comprising words and a tune. When the tune is played, the act of standing implies acceptance of the words (at least, of the first verse), beginning “God save our gracious Queen. . .”. Clearly, God is not just any kind of god. He/She is not one who, having created the Universe, leaves it to run itself ever after; He/She may be sensitive to our supplications and is able to intervene on behalf of a named individual.

So does the national anthem imply a national theology, perhaps that of the established church? If so, some may not want this; the proposed poll could show how many. To avoid controversy, should not the song be renamed the religio-royal anthem?

VERNON R. PICKLES

*65 Yarnells Hill,  
Oxford  
OX2 9BE, UK*

# Research competition with strings

Everyone agrees that support for New Zealand research needs reorganizing, but there is disagreement about the best way of going about it.

## Wellington

NEW Zealand research is in a great tizzy. The Labour government of Mr Geoffrey Palmer is in the throes of radical reorganization of public support for research on the eve of an election (due on 26 October) which, according to the opinion polls, it is likely to lose. There is bipartisan support for the notion that there should be a reorganization, which has been much discussed in the past six years. But the research community fears that the new plans will plunge it from the frying-pan into the fire.

Many of the present anxieties came to the fore at last week's annual meeting of New Zealand chemical societies at which, at one stage, Mrs Margaret Austin, the Minister of Research, Science and Technology, and the opposition spokesman (for the National party), Mr Simon Upton, indulged in the somewhat unconstitutional exercise of commenting on an explanation of the reformed structure offered by her chief executive, Dr Basil Walker. But it remains far from clear what the outcome of the reorganization will be even if the present government wins October's election.

One objective of the reorganization is transparency, which is generally applauded; the government will declare publicly its objectives in research, and will then arrange that the funds available from the public purse will be allocated so as to achieve them. For that purpose, there is (since last December) a Foundation for Research, Science and Technology established as a statutory agency reporting to the minister, but responsible to the New Zealand parliament.

The foundation will eventually allocate the whole public subvention of civil research (estimated to be NZ\$292 million in the financial year from 1 April) to appropriate 'research providers'. The ministry, on the other hand, will be responsible for providing the government as a whole with advice on scientific developments as well as for the definition of research objectives, which Austin promises will be a wide consultative process. (Walker wishes there were more time for the definition of objectives for 1991-92, which must be completed by November.) There is also a cabinet committee (under Austin) for coordinating policy between various ministries.

For researchers, there are several sources of confusion, of which the chief is

the language in which the new policy for research is couched. Research organizations are renamed "providers", the product of research is "output", its cost is "input" and the criterion by which funds will be allocated is known as "contestability". Walker last week, while apologizing for the "economics-based" language, did little to clarify what is meant by "contestability": it is "principally about enhancing and improving the relevance and quality of science in New Zealand . . . focusing on the needs of our society . . . promoting a greater awareness of the necessity of making the very best use of limited resources". That sounds like competition for funds within a predefined framework of objectives. But Walker was at pains to emphasize that the new foundation will have, at any time, a "portfolio" of research investments ranging from short-term applied projects to longer-term (perhaps ten-year) projects as well as projects in basic research where, outputs and inputs would be virtually indistinguishable.

A further difficulty is that New Zealand's seven universities are, for the time being, excluded from the pool for the support of "public good research". Indeed the government is now in the thick of negotiation with vice-chancellors on the conditions under which they would be able to compete for (or "contest") funds available from the pool. It seems reluctantly to have been agreed by the universities that they will have to sacrifice some of their recent support in return for access to a research pool augmented by their own contribution to it. The question is simply how much they should give up. The universities have offered NZ\$5 million, but the government is asking for NZ\$20 million. One university vice-chancellor says that that would cost the universities 400 jobs, about 5 per cent of the total.

Much will hang on how priorities are defined. One danger is that the process of consultation will match the priorities to the present pattern of research, with the result that research organizations will extrapolate their present activities into the future. Among academic researchers, there is great resentment that they have traditionally been disadvantaged, in equipment and research support, compared with researchers at the Department of Scientific and Industrial Research (DSIR), the chief government research agency. But DSIR has already been required to give up "near-market

research", seeking support for that from industry rather than from government. (The chemistry division boasts that it now receives NZ\$22 million from the government and NZ\$14 million from other sources.) Even if the research priorities reflect the status quo accurately, there will be fewer crumbs from DSIR's table to sweep in the direction of academics.

Perhaps a greater danger is that priorities will be too fashionably determined. Hitherto, much government support for research has been directed towards the needs of primary industries such as agriculture, forestry and fisheries. The 50-strong New Zealand Wool Research Organization has, for example, an international standing in its field. Many of these organizations are alarmed that the near-market doctrine and the practice of "contestability" will see a further shrinkage of the proportion of their support that comes from public funds, in most cases already far below the 50 per cent offered when they were established. The risk of that happening will be all the greater if the priority-setting process pays scant attention to the primary industries.

But the poor cousins in the new arrangements, as in the past, are likely to be academic researchers. Not to join the pool will be a recipe for short commons, but to join it may imply that the pattern of academic research will be externally decided. Matters are made even worse by the general economic pressure to which the universities are now subjected. For the first time, undergraduates now pay tuition fees (NZ\$1,200 a year). Research funds out of general university income are sparse — NZ\$1,500 a year for each of a department's PhD students seems quite general. Academics' own access to research funds often depends on collaboration with better heeled partners, perhaps those in government laboratories. But there is a general fear that contestability will mean that research partners are less generous than in the past.

The most curious feature of the reorganization is that it is so self-contained. While the policy-makers insist that a small country (with a population less than that of Scotland) cannot do everything, little attention has as yet been given to collaboration on research projects with partners elsewhere as a means of lessening the risks of bad choices — and of alienating those whose research interests lie outside its current priority-set.

John Maddox



# Galaxies — then you saw them, now you don't

Richard S. Ellis and Carlos S. Frenck

MANY cosmologists believe the Universe has a 'flat' geometry. This is the simplest solution of Einstein's equation and is predicted by fashionable theories of the early Universe. The cosmic expansion generated by the Big Bang would continue indefinitely as the expansion and the gravitationally induced deceleration would be finely balanced. Other solutions to Einstein's equations are possible, however, and the job of the observers is to find the right answer by looking in the sky. Some of the fruits of this endeavour were discussed at a recent meeting\*, where the participants identified a disappearing population of remote galaxies whose fate is likely to be of great cosmological importance.

In the 1920s, the pioneering American astronomer, Edwin Hubble, selected a small representative patch of sky and used photographs taken with the 100-inch Mt Wilson telescope to count galaxies as a function of their apparent brightness. So long as their distribution of intrinsic luminosity is the same in all places and at all times, he argued, the increase in number with decreasing apparent brightness describes how volume varies with depth, so delineating the geometry of the Universe.

It soon became apparent that Hubble's cosmological test is fraught with complications. We see galaxies at large distances as they were a long time ago. Their properties, used in the test merely as tracers of the geometry, could since have changed substantially. Indeed, Beatrice Tinsley, of Yale University, demonstrated in the late 1960s that most kinds of galaxies should have been more luminous in the past and that this would render Hubble's test ineffective, at least for its original purpose.

## Galaxy counting

Motivated instead by galaxy evolution, rather than by cosmic geometry, galaxy counting began as a serious business in the late 1970s. Combining adequate depth and large samples requires panoramic photography with big telescopes. Objectively finding and measuring the brightness of thousands of faint galaxies demands fast machines for measuring photographic plate and sophisticated image processing. More recently, sensitive charge-coupled devices (CCDs) have pushed the counting limit even fainter.

At the meeting, it became apparent that the subject may now be turning full circle and that galaxy counts may, after all, be telling us something about the geometry of our Universe, in addition to providing

new clues as to how galaxies formed and evolved. Groups at Berkeley and Durham Universities and AT&T Bell Labs had already found more blue galaxies at faint limits than would be expected if distant galaxies were identical to those seen nearby. This was originally taken as confirmation of Tinsley's predictions; young galaxies would be more luminous because of massive blue stars. As a result, faint blue galaxies would, on average, be farther away than expected in the absence of evolution and sampling deeper volumes would thus reveal more of them.

Galaxy distances are measured via spectroscopic redshifts — the Doppler frequency change in spectral lines due to the cosmic expansion. Such observations are much harder to make than simply counting. One of the main conclusions presented independently by observing teams at Durham and Cambridge Universities (reported by M. Colless) and the University of Hawaii (L.L. Cowie and S.J. Lilly) is that the bulk of the faint blue population is nearby, contrary to the predictions of the standard picture. Although D.C. Koo (Lick Observatory) warned against rejecting the conventional picture until reliable redshifts are secured for all galaxies in the samples, the incompleteness is small compared to the large count excesses found.

The controversy was fuelled further by new results from a photographic survey of brighter galaxies undertaken using the automated plate measuring (APM) facility at Cambridge. The survey covers a large fraction of the southern sky and yields an accurate measurement of the present volume density of galaxies. W. J. Sutherland (Oxford University) described how the APM survey gives a 'local benchmark' count lower than previously thought, making the relative excess at faint magnitudes even more dramatic. Perhaps, argued T. Shanks (Durham University), the count is depressed locally because our Galaxy lies in a large void whose density of galaxies is abnormally low. Despite recent claims of large-scale cosmic structures, it seems unlikely that the bulk of the southern sky survey is underabundant in galaxies by the required amount. And a hole of the requisite proportions has not turned up in southern redshift surveys.

An interesting solution to the dilemma is offered by T. J. Broadhurst (Queen Mary and Westfield College). Most of the galaxies astronomers study locally are fairly luminous, like our Milky Way. Yet these are just the tip of the iceberg.

Amongst an abundant underlying population of dwarf galaxies, suppose a fraction are somehow triggered into a spectacular burst of star formation which raises their luminosity, rendering them temporarily visible. If such bursts were stronger and/or more common in the past, this could explain the increased galaxy numbers without the attendant increase in average distance. Such a hypothesis could also explain an excess of galaxy counts found by W. Saunders (Oxford University) from the Infrared Astronomical Satellite (IRAS) — a survey instrument sensitive to strongly star-forming galaxies.

## Overabundance

The sheer number of galaxies found — about  $2 \times 10^{10}$  over the whole sky to the current limits of optical detection — already poses a big challenge to conventional ideas. Elementary considerations indicate that, if the co-moving galaxy number density has remained approximately constant in time, then the maximum volume that can be probed at optical wavelengths (before atomic absorption causes the redshifted light from a galaxy to become too dim) is simply insufficient to accommodate the huge number of galaxies counted. Could galaxies have been more abundant in the past and then disappeared in some way? Galaxy mergers, a favoured explanation for origin of elliptical galaxies (see the News and Views article by J.E. Barnes *Nature* **344**, 379; 1990), would have just this effect. A rather large dose of merging would be required however, involving perhaps as much as half of the mass visible today. And, as J.P. Ostriker and P.J.E. Peebles (Princeton University) reiterated, such activity must not be allowed to disturb the fragile stellar disks of spiral galaxies.

Mergers form an integral part of a model of galaxy formation put forward by S.D.M. White (Steward Observatory) in which galaxies form by the cooling and fragmentation of gas onto pre-existing halos made of exotic cold dark matter. This and related models (P. Coles, University of California, Berkeley; C. Lacey, Oxford University; B. Guiderdoni & B. Rocca-Volmerange, Paris University) show that mergers can explain the large number of counts in a flat universe, but it remains to be seen whether they are also consistent with the spectroscopic data. One thing is clear: a large population of hitherto unseen dwarf galaxies is required in these models.

Cowie argued that the reservoir of dwarf galaxies is likely to be insufficient for either bursts or mergers to work and raised the spectre of a Universe with a non-zero cosmological constant,  $\Lambda$ . This constant was introduced by Einstein, before Hubble's discovery of the universal expansion, to maintain a static configuration in universes which would otherwise

\*Galaxies at High Redshift, Oxford University, 30 July–3 August 1990.

collapse by gravity. Although just a constant of integration in his equations, it represents a repulsive force whose introduction Einstein later regarded as the biggest blunder of his career. But as Peebles pointed out, a non-zero  $\Lambda$  produces a healthy increase in the volume available to a given distance, even in a flat universe, and, like it or not, can resolve much of the galaxy-count/redshift dilemma.

Many cosmologists at the meeting shook their heads: adjusting the cosmological geometry to get out of the mess is regarded as the astronomer's last resort. So, is the choice only between a weirder cosmological model than even Einstein wanted and a rather contrived Universe in

which galaxies are miraculously consumed by some cosmic vacuum cleaner? It may well be that the resolution of the faint-galaxies dilemma requires a mixture of more reasonable adjustments: a pinch of large-scale structure, a dab of merging and bursting and maybe even some traditional luminosity evolution. The participants agreed that this would be the less interesting outcome. More exciting is the possibility that the most elementary cosmological test fails because of some important and hitherto undiscovered, property of galaxies and the Universe. □

Richard S. Ellis and Carlos S. Frenck are in the Department of Physics, University of Durham, Durham DH1 3LE, UK.

## GENETICS

# If it smells like a unicorn . . .

F. W. Stahl

In the latest issue of *Genetics*, Barry Hall<sup>1</sup> describes new evidence for environmentally directed point mutations in bacteria. Hall's work extends and strengthens that of Cairns, Overbaugh and Miller<sup>2</sup>, who concluded that a strain of the bacterium *Escherichia coli* unable to use lactose (Lac<sup>-</sup>) had a high rate of reversion to Lac<sup>+</sup> when placed on a lactose medium. Hall's new evidence on point mutations<sup>1</sup> is important because many readers thought that Cairns *et al.* had not adequately supported their conclusions (see, for example, the Scientific Correspondence in *Nature* **336**, 525–528; 1988). Hall's earlier work<sup>3</sup>, which showed that a mutation allowing salicin use by *E. coli* occurred only when salicin was in the medium, was also challenged. His mutation to salicin utilization involved the excision of a transposon, and could be regarded as a special case<sup>4</sup>.

The idea that mutations occur without regard to their immediate usefulness has long been central to our view of evolution. So central, in fact, that both Cairns and Hall must have been reminded what happens to him who strikes the king a non-lethal blow. Although none of the critics convincingly dispatched Cairns's contention, some did give comfort to Defenders of the Faith by pointing out that one aspect or another of the reported work could be variously interpreted<sup>5</sup>. The number and variety of these attacks amplified the admission by Cairns *et al.*<sup>2</sup> that their paper may have been somewhat unpersuasive.

Some commentators missed the point entirely, arguing that the paper by Cairns *et al.*, by a supposed implication of purpose behind directed mutation, challenged the idea of evolution by natural selection (ref. 6, countered in ref. 7). Further, the point of the paper was not that nondividing

bacteria can mutate. That conclusion had been reached earlier<sup>8</sup>, eliminating trivial explanations<sup>9,10</sup>. The point was that the rate of mutations of a beneficial type was higher than the rate of mutation of a type that was adaptively neutral under the conditions of the test.

The neutral phenotype chosen by Cairns *et al.* was resistance to growth inhibition by valine. Because this measurement can be made on almost any *E. coli*, the authors were not inconvenienced by having to construct special strains for their work. But comparing one adaptively relevant mutation (reversion to Lac<sup>+</sup>) to one irrelevant one (mutation to Val<sup>R</sup>) did not seem adequate to some critics. Cairns *et al.* assumed that Lac<sup>+</sup> cells arose in response to energy starvation in the presence of lactose in a special, adaptive way, but the failure of Val<sup>R</sup> mutants to arise was more or less characteristic of mutants in general. With comparable economy, however, the observation could be taken to mean that Val<sup>R</sup> mutants were special, whereas Lac<sup>+</sup> mutants were behaving in the manner typical of most other mutants occurring, but not measured, in the same experiments.

Hall's new contribution<sup>1</sup> not only eliminates that particular possibility but offers interesting new kinds of data. First, Hall generalizes the result of Cairns *et al.* by showing that an amino-acid auxotroph (Trp<sup>-</sup>) will revert to prototrophy on a growth medium lacking tryptophan. On the same medium, Val<sup>R</sup> mutants do not arise. Hall then shows that a Cys<sup>-</sup> mutant reverts on a medium lacking cysteine. The double auxotrophic strain, Trp<sup>-</sup>Cys<sup>-</sup>, provided the material for the critical test of specificity. The double mutant reverts only to Trp<sup>+</sup> when the culture is starved of tryptophan and only to Cys<sup>+</sup> when starved of cysteine. Thus, each mutant character

reverts when the bacteria can profit from the reversion, but neither character reverts when the reversion would confer no immediate growth advantage.

Cairns, too, has done new experiments that strengthen the argument for directed mutation. An unfortunate feature of the paper of Cairns *et al.*<sup>2</sup> was the wimpy mutational response of the *lac* amber strain used in the studies. Now Foster and Cairns have repeated those studies using frameshift mutations in *lac* (personal communication). Some of these mutant alleles revert like gangbusters on medium containing lactose as the only energy source, while mutating at a rate too low to measure accurately when the cells are grown on other energy sources. This robust adaptive mutational response has facilitated the application of more direct methods of measurement, allowing Foster and Cairns to exclude even the most perverse of alternative interpretations. Neither Hall nor Foster and Cairns have demonstrated a molecular mechanism for environmentally directed mutagenesis. But both sets of authors have promised to test extant models, one of which is proffered by Hall<sup>1</sup>, in the near future.

In the distant past, Ryan<sup>8-10</sup> reported reversion of a His<sup>-</sup> mutant starving for histidine. (Those mutants were conveniently detected as colonies arising days after a His<sup>-</sup> culture was plated on medium lacking histidine.) Had Ryan looked at some adaptively irrelevant mutant type, instead of His<sup>+</sup>, in the histidine-starved culture, he would have seen that non-dividing cells do not mutate at a rate detectable by his methods. Had he looked both at His<sup>-</sup> and at this other mutant type, he might have recalled Max Delbrück's warning of 1946, which was delivered to André Lwoff at a symposium attended by Ryan<sup>11</sup>: "In the case of mutations of bacteria . . . to phage resistance . . . the phage does not cause the mutations. In your case of mutations permitting the mutants to utilize succinate . . . as sole carbon and energy source . . . it [is] an obvious question to ask whether this particular medium had an influence on the mutation rate . . . One should keep in mind the possible occurrence of specifically induced adaptive mutations." □

F.W. Stahl is in the Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403-1229, USA.

- Hall, B.G. *Genetics* **126**, 5–16 (1990).
- Cairns, J. *et al.* *Nature* **335**, 142–145 (1988).
- Hall, B.G. *Genetics* **120**, 887–897 (1988).
- Symonds, N. *Nature* **337**, 119–120 (1989).
- Lenski, R.E., Slatkin, M. & Ayala, F.J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2775–2778 (1989).
- Rees-Mogg, W. *Independent*, 13 Sept. 1988.
- Maynard Smith, J. *Nature* **336**, 107–108 (1988).
- Ryan, F. *Proc. int. Genet. Symp.* **1956**, 555–558 (1957).
- Ryan, F.J. *J. gen. Microbiol.* **21**, 530–549 (1959).
- Ryan, F.J., Nakada, D. & Schneider, M.J. *Z. Vererb. u. Evol.* **92**, 38–41 (1961).
- Delbrück, M. *Cold Spring Harb. Symp. quant. Biol.* **11**, 139–155 (1947).



# Dating old gold deposits

François Robert

GOLD'S importance to the world's economy is undeniable. It is produced as a primary commodity or as a by-product from many types of ore deposits. Auriferous quartz veins in Archaean rocks (over 2,500 million years old) have historically been a significant source of gold<sup>1,2</sup>. But new deposits are increasingly hard to find and elucidating the geological processes leading to the formation of these gold-quartz vein deposits is a necessary step in improving exploration strategies. Various models of the processes have been proposed, but a key element in choosing between them is a precise knowledge of the deposits' ages relative to the geological processes inferred to be responsible for their formation. Difficulties in obtaining good ages have hampered progress, but on page 831 of this issue<sup>3</sup>, Jemielita *et al.* report one of the first studies to apply high-precision U-Pb dating techniques to Archaean gold mineralization and associated host rocks. Their data from the Camflo gold deposit in the Abitibi greenstone belt of Canada corroborate the results of others and lead to some surprising conclusions which require a re-evaluation of current concepts.

Archaean gold-quartz vein deposits typically consist of networks of quartz-carbonate veins (see figure) occurring almost exclusively within belts of deformed and metamorphosed volcanic rocks, known as greenstone belts. Within these belts, the auriferous veins are distributed along large, crustal-scale faults and are generally associated with smaller, subsidiary faults<sup>1,2</sup>. The veins occur in all rock types in greenstone belts but show a common spatial association with small intrusions of felsic composition (diorite, granodiorite, monzonite). There is a general agreement that gold-quartz veins were emplaced relatively late in the evolution of their host greenstone belts, after or during the main phase of mountain building, or orogenic, deformation<sup>1,2</sup>. The deformation involved folding, faulting and metamorphism of volcanic and associated rocks; it was accompanied and slightly followed by the emplacement of intrusions, including the mineralized felsic intrusions.

The nature of the faults hosting the gold-quartz veins and the overall metamorphic grade of the greenstone belts suggest that these deposits, now exposed at the Earth's surface, formed originally at

depths of the order of 10 km or more<sup>4</sup>. It is generally agreed that the auriferous H<sub>2</sub>O-CO<sub>2</sub> fluids that produced the deposits were generated at deeper portions of the crust and migrated up along major fault zones and other favourable structures<sup>2,4,5</sup>. But there is no consensus on the origin of these auriferous fluids. The magmatic model proposes that, at least for some deposits, the fluids and the gold are derived from the same crystallizing magma as the host felsic intrusions<sup>6,7</sup>. According to the metamorphic-



Photograph of the complex intersection between steep and horizontal auriferous quartz (white)—tourmaline (black) veins from the Sigma mine, Val d'Or, Canada. The arrow indicates rocks containing hydrothermal rutile analogous to that dated at the Sigma and Lamaque deposits<sup>12,17</sup>. (Scale bar, 1 m.)

replacement model<sup>1</sup>, the fluids are produced by prograde metamorphism in the deeper portions of the greenstone belts, which are also the ultimate source of gold. Other models relate the generation of fluids to deep crustal processes involving mantle CO<sub>2</sub> outgassing and granulitization of the lower crust, which would also be accompanied by magmatism and metamorphism in the lower crust<sup>5,8,9</sup>. In all of these models, the emplacement of gold deposits is inferred to be closely related to the late stages of greenstone belt orogenesis.

In addition to having to account for constraints imposed by various kinds of geological and geochemical data, any model for the source of gold-bearing fluids is valid only if there is a temporal coincidence between the gold mineralization and the proposed fluid-generating processes. Recent technical refinements have allowed scientists to obtain high-precision U-Pb ages from zircons in volcanic and plutonic rocks, providing a framework for the evolution of many greenstone belts. In the Abitibi greenstone belt, such work has shown that volcanism ended about 2,700 million years (Myr) ago and that magmatism generating the intrusions continued until about 2,673

Myr ago<sup>10-12</sup>. The age of metamorphism, which also marks the last stages of orogenesis, is less well constrained but has been estimated<sup>13</sup> at 2,680 Myr.

Early attempts to date the gold deposits in the Abitibi belt were based on <sup>40</sup>Ar-<sup>39</sup>Ar geochronology of hydrothermal micas that presumably precipitated from the same fluids as the gold. These studies yielded ages of 2,633 ± 6 and 2,579 ± 3 Myr at the Dome<sup>14</sup> and Sigma<sup>15</sup> deposits, respectively, suggesting that the gold could be significantly younger than both the intrusions and the metamorphism observed at the exposed crustal level. But, as pointed out by Jemielita *et al.* in this issue<sup>3</sup>, these young ages could not unequivocally be interpreted as representing the

age of gold mineralization because they could reflect the age of later cooling and uplift of the greenstone belt.

The new work, on the Camflo deposit where gold mineralization is hosted by a small felsic intrusion, represents the first published detailed study involving high-precision U-Pb dating of minerals considered to be less susceptible to later geological events such as cooling and uplift. It follows directly on similar work at Camflo by P. L. Zweng<sup>15</sup> who, through unpublished detailed underground mapping, provided the essential geological basis for new geochronological

work. Jemielita *et al.*<sup>3</sup> obtain an age of 2,680 ± 4 Myr for the host intrusion from magmatic zircon and titanite, and an age of 2,625 ± 7 Myr for the gold mineralization from hydrothermal titanite and rutile associated with gold. The implication of these ages is that the gold mineralization is not related to magmatism or metamorphism at the time of orogenesis as proposed in some models<sup>2,6,7</sup>.

Additional geochronological studies on other gold-quartz veins from the same area as the Camflo deposit also yielded similar ages. Scheelite from three deposits of the Val d'Or area give a Sm-Nd age<sup>16</sup> of 2,602 ± 20 Myr. U-Pb dating<sup>3,12,15,17</sup> of hydrothermal rutile and titanite at the Sigma, Lamaque, Camflo and Kerr Addison deposits give ages ranging from 2,577 ± 3 Myr to 2,625 ± 7 Myr. In all these deposits, the gold-quartz veins are thus separated from magmatism and metamorphism (at the exposed crustal level) by at least 55 million years. The ages obtained from different hydrothermal minerals using different isotopic systems are thus broadly similar. Such consistency supports the interpretation that these young ages represent the age of gold mineralization rather than the age of later geological events such as cooling

and uplift, which would be expected to be different for different minerals and isotopic systems.

Therefore, geologists will have to consider deeper crustal processes as those responsible for generation of the auriferous fluids, as proposed by some models<sup>5,8,9</sup>. As pointed out by Jemielita *et al.*<sup>3</sup>, such models are supported by the fact that similarly young ages have been obtained from an area of high-grade metamorphism believed to represent currently exposed portions of the lower crust under the Abitibi belt. These young ages indicate that metamorphic and magmatic activities in the lower crust continued long after those at mid- to upper crustal level, and that gold-quartz veins at mid- to upper crustal levels may be a manifestation of continued deeper crustal processes, as initially proposed by Corfu<sup>18</sup>.

As is generally the case when new and perhaps controversial data or ideas are presented, there will probably be a debate on the validity and significance of the new measurements. For example, recent ion-probe U-Pb dating of zircons from the same Val d'Or deposits<sup>19</sup>, interpreted as hydrothermal minerals related to gold on the basis of rather limited textural evidence, yielded significantly older ages than the new studies. These ages are in fact close to the age of the respective host intrusions. But the real importance of geochronological work such as that presented by Jemielita *et al.* is that it forces those who study gold-quartz vein deposits to ask themselves new questions and to consider new options for the origin of these deposits. □

François Robert is at the Geological Survey of Canada, 601 Booth St, Ottawa, K1A 0E8, Canada.

## Peptide feeding and cellular cookery

Peter Parham

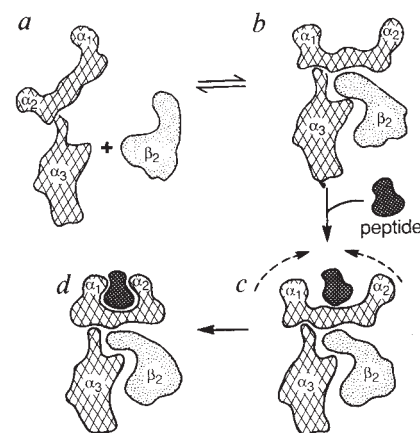
OVER the past few years, the role of major histocompatibility complex (MHC) glycoproteins in antigen presentation to T lymphocytes has become established. Peptides derived from 'self' and 'foreign' proteins bind intracellularly to MHC molecules and are transported to the cell surface where the MHC peptide complex can be recognized by T-cell receptors.

Empty H-2 molecules and their interactions with peptides. Ljunggren *et al.*<sup>6</sup> *in vivo* and Townsend *et al.*<sup>8</sup> *in vitro* show that H-2 heavy chains can associate (a) with  $\beta_2m$  to give heterodimers with a conformation that has an operationally empty binding groove. (Jack Strominger has pointed out that it probably contains water.) Molecules in this open state (b) rapidly bind appropriate peptides (c), forming stable complexes — the closed state — in which the peptide binding site is virtually, but not quite, inaccessible<sup>11</sup> (d). Schumacher *et al.*<sup>7</sup> propose that the transition from open to closed state involves a peptide-induced conformational change and thereby bring relish to the familiar hot-dog-in-a-bun model of peptide binding. Furthermore, this model may explain the paradox that a goodly proportion (up to 25 per cent) of purified class I molecules bind solid phase peptides<sup>14</sup> whereas less than 1 per cent bind to the same peptide in solution<sup>11</sup>. Interactions occurring at the solid phase surface might facilitate transition from the closed to the open state and thus allow an exchange of bound peptide. In their analysis of *in vitro* assembly in detergent lysates of whole cells, Townsend *et al.* find that addition of either peptide or  $\beta_2m$  will promote the formation of class I molecules that can be recognized by specific antibodies. They propose a more general scheme for class I assembly (see Fig. 6 in ref. 8) involving equilibria between free components, complexes consisting of heavy chain with just  $\beta_2m$  (empty molecules) or just peptide and mature complexes consisting of all three components<sup>8</sup>.

More recently, it has emerged that bound peptide is important for the stability, conformation and cell-surface expression at least of the class I MHC molecules. Contributing to this latter conclusion have been studies on RMA-S, Klas Kärre's mutant murine cell line<sup>1</sup>, which is defective in class I MHC expression despite normal genes encoding H-2 heavy chains and  $\beta_2m$ -microglobulin ( $\beta_2m$ ).

In particular, Townsend *et al.*<sup>2</sup> reported in *Nature* last year that surface expression of class I molecules in RMA-S cells could be increased by 'feeding'<sup>15</sup> the cells with specific peptides, and analysis of biosynthetically labelled class I molecules showed enhanced association of heavy chain with  $\beta_2m$ . Subsequent studies have confirmed this effect for other cells and class I molecules<sup>4,5</sup>. The interpretation of these results favoured by Townsend *et al.*, is that by occupying the peptide binding site on the class I heavy chain, peptide stabilizes its folding and association with  $\beta_2m$ . They suggest that at the high peptide concentrations with which the cells are 'fed', some peptide may reach a pre-Golgi compartment of the cell and there

stimulate the assembly of newly synthesized class I molecules, enabling them to proceed through the Golgi to the cell surface. Recent papers in *Nature*<sup>6</sup> and *Cell*<sup>7,8</sup> however show that class I assembly and cell-surface expression can occur in the absence of added peptides; and that the effects of extracellular peptides on class I assembly and expression in RMA-S



cells can be explained without invoking retrograde transport to the endoplasmic reticulum (ER). All of these papers confirm the central importance of peptide for class I assembly, conformation and structural integrity<sup>6-8</sup>. But exactly how and where the peptide makes its contribution is not yet clear. What the new papers do illuminate is the question of whether heterodimers of class I heavy chains and  $\beta_2m$  can ever exist without bound peptide, as what David Margulies has termed 'naked' MHC molecules<sup>9</sup>.

This question arose directly from the crystal structure of the class I molecule which was shown to contain bound peptide<sup>10</sup>; and from the inefficiency with which added peptides can be bound to purified class I molecules<sup>11</sup>. What Hidde Ploegh and colleagues have now shown, by some innovative 'cellular cookery'<sup>6,7</sup> is that class I molecules will accumulate at the surface of RMA-S cells in the absence of added peptide if the temperature is right. RMA-S cells cultured at 37 °C produce rapidly degraded H-2 heavy chains, most of which neither acquire sialic acid nor reach the cell surface.

1. Roberts, R.G. *Geosci. Canada* **14**, 37–52 (1987).
2. Groves, D.I. & Phillips, G.N. *Ore Geol. Rev.* **2**, 287–322 (1987).
3. Jemielita, R.A., Davis, D.W. & Krogh, T.E. *Nature* **346**, 831–834 (1990).
4. Sibson, R.H., Robert, F. & Poulsen, K.H. *Geology* **16**, 551–555 (1988).
5. Cameron, E.M. *Geology* **16**, 109–112 (1988).
6. Burrows, D.R. & Spooner, E.T.C. *Econ. Geol. Monograph* **6**, 424–444 (1989).
7. Cameron, E.M. & Hattori, K. *Econ. Geol.* **82**, 1177–1191 (1987).
8. Fyon, J.A., Troop, D.G., Marmont, S. & Macdonald, A.J. *Econ. Geol. Monograph* **6**, 479–490 (1989).
9. Perring, C.S. *et al. Econ. Geol., Monograph* **6**, 571–584 (1989).
10. Corfu, F., Krogh, T.E., Kwok, K. & Jensen, L.S. *Can. J. Earth Sci.* **26**, 1747–1763 (1989).
11. Mortensen, J.K. *Geol. Surv. Canada, Paper* 87–1A, 581–590 (1987).
12. Wong, L. *et al. F. Abs. Geol. Assoc. Canada–Mineral. Assoc. Canada Meeting* **14**, A45 (1989).
13. Hanes, J.A., Archibald, D.A., Hodgson, C.J. & Robert, F. *Geol. Surv. Canada, Paper* 89–1C, 135–142 (1989).
14. Masliwec, A., McMaster, D. & York, D. in *Ontario Geol. Surv. Misc. Paper* 130 (ed. Mine, V.G.), 107–114 (1986).
15. Zweng, P.L. & Mortensen, J.K. *Abstr. Geol. Soc. Am. A Meeting* A351 (1989).
16. Anglin, C.D. *Geol. Surv. Canada, Paper* 90–1C, 255–259 (1990).
17. Jemielita, R.A., Davis, D.W. & Krogh, T.E. *Abstr. Geol. Soc. Am. A Meeting* A351 (1989).
18. Corfu, F. *Precambrian Res.* **36**, 259–275 (1987).
19. Cloué, J.C., King, R.W. & Kerrich, R. *Earth planet. Sci. Letters* **98**, 109–128 (1990).



'Cooking' at lower temperatures — 26 °C was optimal — produced an accumulation of mature, terminally glycosylated molecules at the plasma membrane. By structural and functional criteria, these molecules behave like no other class I MHC molecules previously studied and have properties appropriate for unstable, peptide-free dimers of heavy chain and  $\beta$ 2m (see table). Ljunggren *et al.* call them "empty class I molecules that come out in the cold"<sup>1</sup>.

But empty class I molecules also seem to come out in 'the hot' (or at any rate at 37 °C); although like Leamas, the spy who came in from the cold in Le Carré's novel<sup>2</sup>, they do not survive long on arrival. At 37 °C, most H-2 heavy chains in RMA-S cells never reach the plasma membrane, but the results of Ljunggren *et al.* indicate that a small proportion continuously arrive as empty molecules, which then rapidly fall apart. Empty molecules can, however, be stabilized by the binding of an exogenously added peptide<sup>3</sup> (see figure), and also by the binding of specific anti-H-2 monoclonal antibodies (G. Hämmerling and V. Ortiz-Navarrete, personal communication). The unexpected observation that empty class I molecules reach the cell surface and can be rescued there by peptide binding suggests an alternative and less radical interpretation to that favoured by Townsend *et al.* from the original peptide-feeding experiments<sup>2</sup>.

Their interpretation was that peptides use an unusual pathway of intracellular transport to enter cells and facilitate class-I assembly in the endoplasmic reticulum, the alternative merely requires that they have access to the cell surface. Although the first interpretation is not excluded, the simpler alternative can reasonably explain the existing data. Prolonged incubation with peptide — as in the original feeding experiments<sup>2</sup> — would be expected to give rise to an accumulation of stabilized molecules, and hence to enhanced expression of antigenically recognizable class I molecules. The high peptide concentration required for this effect to be seen is clearly not necessitated by a low affinity of empty MHC molecules for peptide — Schumacher *et al.*<sup>7</sup> show that binding is rapid and, once bound, peptide does not dissociate — but probably because the lifetime of an empty class I molecule at the cell surface is so short. At lower temperatures they can be induced to survive for longer, in which case less peptide is required<sup>6,7</sup>.

Thus there are now two ways to increase class I expression at the RMA-S cell surface: lowering the temperature of culture, and incubation with specific peptides. Ljunggren *et al.*<sup>6</sup> find that the two treatments are synergistic, indicating that peptide is more effective at surface stabilization, but that lower temperature increases the rate at which empty class I

molecules successfully navigate the exocytic pathway.

From these new results it seems that enhanced class I expression in peptide-fed RMA-S cells can be explained by cell-surface stabilization of a flux of empty H-2 molecules, involving a small proportion of total H-2 heavy chains synthesized by the cells. How then does one explain the following observations of Townsend *et al.*<sup>2</sup>: first that in extracts from cells both peptide-fed and biosynthetically radiolabelled a significant proportion of heavy chains were induced to assemble with  $\beta$ 2m, and second, that these heavy chains are sensitive to endoglycosidase-H and therefore should derive from an intracellular, pre-Golgi compartment? Under the conditions used in those experiments, a long incubation with peptide followed by a shorter radiolabelling period, it is likely that most empty H-2 molecules that actually bound peptides at the surface of intact cells, would not have been radioactive and would thus have been 'invisible'

to immunoprecipitation, electrophoresis and autoradiography. A clue that the observed association of newly synthesized polypeptides in peptide-fed cells might not result from an *in vivo* process was the fact that almost all  $\beta$ 2m-associated heavy chains remained in immature endoglycosidase H-sensitive forms. This suggested that they were not moving from the Golgi complex to the plasma membrane as shown for class-I molecules *in vivo*<sup>13</sup>.

In extending and re-evaluating these observations, Townsend *et al.*<sup>8</sup> now show the observed assembly can be accounted for by reactions occurring after disruption of the cells with detergent. They find that class-I assembly can in fact be induced by addition of peptide to extracts made from unfed RMA-S cells and similar results were seen on addition of extra  $\beta$ 2m, providing independent evidence for the existence of empty molecules. A key experiment was to mix detergent lysates from two populations of RMA-S cells — one peptide fed, the other radiolabelled

## A head in the sands of time

WILD ostriches in naturalistic poses are among the most common motifs of the rock engravings of the Sahara. The numerous rock-art sites in the valley of the Enneri Blaka, Niger, are no exception: ostriches are depicted standing alone, or in groups, and sometimes as the quarry of hunters. A prehistoric engraving of what appears to be a domesticated ostrich has now been discovered<sup>1</sup>, confounding the view that these birds were domesticated only in historical times.

In the past, one very schematic figure of an ostrich with a man on its back and holding its head had been claimed to represent domestic use of the species<sup>2</sup>: other birds appear to bear pack-saddles just like those in pictures of cattle from the same period. But such interpretations have not generally been taken seriously for the ostrich<sup>3</sup>.

The new discovery is a small, deeply engraved figure, attributed to the 'bovidian' style, and hence to the Neolithic period between about 7,000 and 5,000 years ago (see figure, after E. Tillet: scale bar, 10 cm). Located in rocks on the left bank of the River Blaka, it shows an ostrich in left profile. The wings and feathers are not shown, and the legs are folded beneath the reclining body. The lines above have been interpreted<sup>1</sup> as some kind of load, tied up with a series of

straps, and perhaps fixed to the ostrich at the neck and rump.

In Africa, much use was made throughout prehistory of ostrich feathers for decoration, and of its eggshells for vessels and beads, but it was believed that

there was no ancient attempt at domestication. It is known that the Egyptians occasionally caught young ostriches and broke them to the harness: in the Graeco-Roman period, eight ostriches drew the ceremonial chariot of King Ptolemy Philadelphus, and his queen sometimes rode on one<sup>4</sup>. But true domestication

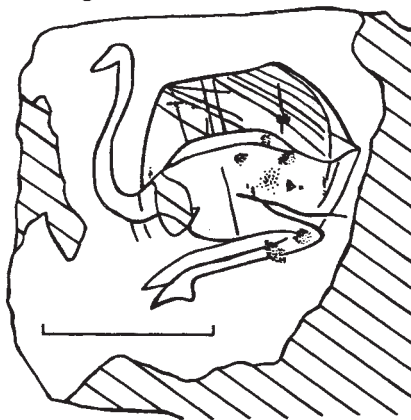
has often been attributed to French experiments in North Africa in the late nineteenth century.

The new discovery, together with the other figures considered doubtful or enigmatic until now, may prompt a re-evaluation of the role of this useful bird in early times.

Paul Bahn

Paul Bahn is a freelance writer on archaeology based at 428 Anlaby Road, Hull HU3 6QP, UK.

1. Tillet, T. *Bull. Soc. prehist. Fr.* 87, 139–140 (1990).
2. Lhote, H. *Les Gravures de l'oued Mammanet* (Nouvelles Editions Africaines, Dakar, 1979).
3. Monod, T. *L'Adrar Ahnet* (Institut d'Ethnologie de l'Université de Paris, 1932).
4. Hyams, E. *Animals in the Service of Man* (Dent, London, 1972).



— and show association similar to that found when the same cells were fed and labelled<sup>8</sup>. Thus the primary role of peptide feeding in such biosynthetic experiments is probably to load the vesicles and organelles of the endocytic pathway with peptides. On detergent solubilization the contents of all intracellular membrane compartments become mixed; peptides from endocytic vesicles then mingle with components of the endoplasmic reticulum and Golgi complex — inducing assembly of free heavy chains with  $\beta 2m$ , stabilizing any empty H-2 molecules<sup>8</sup>.

The recent papers<sup>6-8</sup> show that at least two effects have been at work in experiments examining the H-2 molecules of peptide-fed RMA-S cells: one is peptide binding by empty class I molecules at the cell surface to give enhanced class I expression in whole-cell binding assays<sup>6,7</sup>, the other peptide-driven assembly and stabilization of H-2 heavy chains and  $\beta 2m$  *in vitro*, in the solubilized cell extracts made during immunoprecipitation analysis<sup>8</sup>. These mechanisms for peptide binding do not necessarily conflict with that proposed<sup>2,3</sup> to operate in living cells — retrograde transport of fed peptides to the endoplasmic reticulum — and it is still possible that retrograde transport contributes to the phenomenon although the data no longer demand this interpretation. Unchanged, however, is our understanding of the lesion in intact cells of the RMA-S mutant, which clearly incapacitates assembly of peptides  $\beta 2m$  and class I heavy chains in the endoplasmic reticulum<sup>2</sup>. That added peptide facilitates assembly of class I molecules *in vitro* in RMA-S extracts further supports the original interpretation of Townsend *et al.*<sup>2</sup> that the defect lies with a specialized mechanism that transports endogenously

produced peptides from the cytoplasm and other subcellular compartments across membranes into the lumen of the endoplasmic reticulum. Moreover, the more recent analysis by Townsend *et al.*<sup>8</sup> points to the insights obtained with RMA-S having direct relevance for the results obtained from feeding peptides to other mutant and normal cells<sup>4,5</sup>.

An important practical question that emerges from these studies is whether loading of empty class I molecules contributes significantly to peptide sensitization of target cells in T-cell assays. The observations of Kärre, Ploegh, Townsend and colleagues<sup>6-8</sup> suggest that such effects could be significant, but dependent upon the temperature and the handling of target cells before and during incubation with peptide. In addition to identifying empty class I molecules and the mechanisms of peptide feeding, these studies provide better methods for binding peptides to H-2 molecules (see discussion<sup>7,8</sup>) and insight into the workings of antigen presentation. In the mutant cell, RMA-S, assembly of H-2 heavy chains and  $\beta 2m$  appears severely limited by insufficient peptides. To a lesser but significant extent, the same seems true of RMA (the control)

and other normal cells. The results of Ljunggren *et al.*<sup>6</sup> indicate that up to 15 per cent of class I molecules reaching the RMA plasma membrane are empty, implying that peptide shortage is a fact of healthy life. Teleologically, this makes sense. The antigen-presentation system is hungry, partially starved for peptide; thus poised, it has greater potential to present products associated with viral infection, malignancy and other cellular trauma. □

Peter Parham is in the Beckman Laboratories for Structural Biology, Department of Cell Biology, Stanford University School of Medicine, Stanford, California 94305, USA.

1. Ljunggren, H.G. & Kärre, K. *J. exp. Med.* **162**, 1745–1759 (1985).
2. Townsend, A., Ohlen, C., Bastin, J., Ljunggren, H.G., Foster, L. & Kärre, K. *Nature* **340**, 443–448 (1989).
3. Parham, P. *Nature* **340**, 426–428 (1989).
4. Lie, W.-R. *et al.* *Nature* **344**, 439–441 (1990).
5. Cerundolo, V. *et al.* *Nature* **345**, 449–452 (1990).
6. Ljunggren, H.-G. *et al.* *Nature* **346**, 476–480 (1990).
7. Schumacher, T.N.M., Heemels, M.-T., Neefjes, J.J., Kast, W.M. & Ploegh, H.L. *Cell* **62**, 563–567 (1990).
8. Townsend, A. *et al.* *Cell* **62**, 285–295 (1990).
9. Margulies, D. *Nature* **342**, 124–125 (1989).
10. Bjorkman, P.J. *et al.* *Nature* **329**, 506–512 (1987).
11. Chen, B.P. & Parham, P. *Nature* **337**, 743–745 (1989).
12. Le Carré, J. *The Spy Who Came in From the Cold* (Gollancz, London, 1963).
13. Krangel, M.S., Orr, H.T. & Strominger, J.L. *Cell* **18**, 979–991 (1979).
14. Bouillot, M. *et al.* *Nature* **339**, 473–475 (1989).

#### PALAEOCLIMATE

## Dynamic former ice sheets

Julian A. Dowdeswell

LARGE ice sheets have extended into the Northern Hemisphere mid-latitudes several times during the past two million years in association with changes of climate. However, they did not simply respond to climatic shifts. As they grew over significant parts of North America and Northern Europe, over periods of  $10^3$ – $10^4$  years, the ice sheets in turn modified both climate and sea level. On page 813 of this issue<sup>1</sup>, Boulton and Clark use satellite-derived broad-scale patterns of glacial sediment ridges to determine the way the North American Laurentide ice sheet flowed. Large changes of 1,000–2,000 km in the location of centres of ice distribution and the western margins of the Laurentide ice sheet over approximately the past 120,000 years are indicated by superimposed sets of ridges. These shifts in the centres of mass are argued to be a result of strong coupling and feedbacks between the ice sheet and atmosphere whereby, for example, changing ice-sheet morphology affects the mass-balance function on the ice-sheet surface.

Coupled, three-dimensional numerical models are being used increasingly to investigate and predict the interplay between ice-sheet growth and decay, climate variations and sea-level change. The time-dependent nature of ice-sheet–

climate interactions can be investigated, and models usually include consideration of ice-flow parameters to predict changing ice-sheet geometry and of the Earth's rheology to calculate the effects of ice loading on surface elevation<sup>2-5</sup>. Global changes in sea level provide an external forcing mechanism which influences individual ice sheets, yet variations in regional sea level are driven, in turn, by changes in the thickness of the ice. As ice and snow extend progressively over large areas, surface albedo is raised, reducing energy inputs to the surface and thus allowing ice to accumulate faster.

The relationship between precipitation and changing ice-sheet topography during ice sheet growth is also of particular significance<sup>1</sup>. The relief associated with a continental ice sheet over North America, for example, affects tropospheric wind flow, splitting the mid-latitude Westerlies and modifying temperature and precipitation patterns over the ice sheet and adjacent areas<sup>6</sup>. These changes would presumably lead to further shifts in ice sheet configuration. The specification of atmospheric moisture models and their response to the effects of ice-sheet topography is therefore important<sup>7</sup>.

The model treatment of many important processes remains simplified, however. For example, the basal boundary

#### PROPERTIES OF EMPTY H-2 MOLECULES

1. They have class I heavy chains associated with  $\beta 2m$ .
2. They can reach the cell surface.
3. They are unstable compared with H-2 molecules 'filled' with endogenous peptides and rapidly dissociate/degrade at the cell surface.
4. They bind antibodies that also react with 'full' H-2 molecules.
5. They are more stable at temperatures of 23–31 °C than at 37 °C and steady state levels at cell surfaces are optimized by short-term cell culture at 26 °C.
6. Their structure is stabilized by the binding of peptide.
7. Binding of peptide to empty molecules associated with conformational change.
8. They do not present an endogenous minor histocompatibility antigen to cytotoxic T cells.
9. They bind, and present to T cells, exogenously added peptides more efficiently than 'full' molecules.
10. They probably constitute a proportion of class I molecules on many cell types.



conditions for fast-flowing ice streams and outlet glaciers, basal melting of ice shelves and the mechanisms and rates of iceberg production are often poorly specified. Continuing field and laboratory investigations are needed if the details of boundary conditions are to keep pace with the growing computational sophistication of numerical models.

Not only the dynamics of ice sheets, but also the cause of their inception, have been studied using computer modelling. In particular, the role of the variation of solar heating of the Earth as its orbit varied (the 'Milankovitch' cycles) has been tested using global circulation models with complex atmospheric treatments<sup>7</sup>. The climate and sea-level records from deep ocean sediments vary with a periodicity similar to the Milankovitch parameters<sup>8</sup> including orbital eccentricity and tilt of the Earth's rotational axis. The implication is that these variations resulted in ice sheet growth and decay. Recent computer experiments have failed to show even that snow cover is maintained in response to reduced radiative inputs in summer. This could indicate either that the models are not nearly sensitive enough to climate forcing or that Milankovitch variations alone are insufficient as the direct external forcing factor for initiating the growth of ice sheets<sup>7</sup>. Global circulation models can be used only for climatic snapshots, where all or many of the lower boundary conditions (such as ice-sheet size, sea-ice extent, snow cover, albedo and even sea surface temperature) are prescribed<sup>9</sup>. Modelling the time-dependent growth of ice sheets over thousands of years using this approach is computationally impractical at present<sup>7,10</sup>.

The geological record and numerical modelling are complementary tools in looking at ice-climate interactions. Geological evidence can be used both to test the predictions of numerical ice-sheet models, and to generate hypotheses concerning ice-sheet behaviour, whose physical plausibility can be tested through further modelling. For example, several numerical models of the Antarctic ice

sheet predict that grounded ice advances over the surrounding continental shelves during low, full-glacial sea-level stands<sup>2,3</sup>. This can be tested by reference to the stratigraphy and sediments in cores from these shelves<sup>11,12</sup>. Equally, the rapid disintegration of the Barents Sea ice sheet 15,000 years ago, indicated in the oxygen-isotope record from the Norwegian-Greenland Sea, resulted, according to one hypothesis<sup>13</sup>, from instability brought about by its own mass: as the ice sheet grew, it may have depressed parts of its own bed to a point where they became buoyant and unstable without the requirement for any external climatic forcing to initiate decay<sup>13</sup>. This hypothesis can be tested using a coupled, three-dimensional model including calculations on ice flow and crustal response to ice loading.

In the same way, reconstructing and understanding the changing configuration

of ice sheets require both numerical calculations and field studies. Modelling of former ice sheets, including the Laurentide, depends on a three-dimensional, time-dependent approach, in which ice, ocean, atmosphere and solid Earth are coupled. The contribution of Boulton and Clark<sup>1</sup> is to indicate the dynamic nature of the Laurentide ice sheet through their analysis of flow lines indicated by kilometre-scale ridges. If we can demonstrate, by reference to the geological record, that the dynamics of former ice sheets can be modelled with some success, then we may also approach problems concerning the future effects of possible global warming on modern ice sheets and, hence, sea level with increased confidence. □

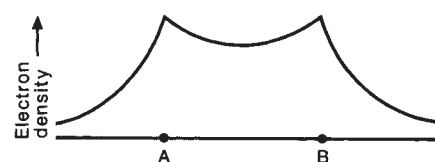
Julian A. Dowdeswell is at the Scott Polar Research Institute, University of Cambridge, Cambridge CB2 1ER, UK.

## CHEMICAL PHYSICS

# A molecular catastrophe

David L. Cooper

CATASTROPHE theory is that curious branch of mathematics that deals directly with situations for which very small perturbations can lead to very dramatic and sudden changes — remember the straw that broke the camel's back. Linked to algebra, calculus and topology, catastrophe theory, is usually associated with the work of the French mathematician



Typical variation of electron density along the path joining two chemically bonded atoms, A and B, showing the two nuclear maxima.

René Thom<sup>1</sup> in the 1960s and 1970s. It has since been applied to a vast range of problems<sup>2</sup> drawn from the physical sciences, engineering and biological processes, and extending to the modelling of weather patterns and even of social systems. In new work<sup>3</sup>, Cioslowski has applied these techniques to a long-standing problem in chemical bonding and, in so doing, he has reinforced a growing realization that all is not well with our understanding of the bonding in supposedly simple systems such as small clusters of metal atoms and bulk metals.

There are seven distinct types of elementary catastrophe, known as fold, cusp, swallowtail, butterfly, elliptic, hyperbolic and parabolic catastrophes. A simple example of a fold catastrophe is provided by the function  $y = x^4 + bx^2$ , where all quantities are real. If  $b$  is positive, then the function possesses a minimum at  $x = 0$ . If

$b$  is negative, then the function exhibits a pair of minima symmetrically placed on either side of a maximum at  $x = 0$ . These two very different regimes are separated by the 'bifurcation point', for which  $b = 0$ . Close to this point, very small changes in  $b$  can lead to very dramatic changes in any physical system to which the equation applies.

In his new study, Cioslowski looks at the electron density of molecular systems, which is related to the probability of finding an electron in a given portion of the molecule. For two atoms chemically bonded to one another, this electron density typically behaves in the fashion illustrated in the figure, with a minimum along the path joining the two atoms. The critical points in the distribution are the locations at which derivatives of the electron density are zero. The positively charged atomic nuclei attract the negatively charged electrons and, in general, the characteristics of electron density maxima occur only at the nuclear positions. In addition, of course, the electrons repel one another strongly.

Small clusters of lithium and of sodium atoms are startlingly different<sup>4-6</sup>: the critical points include maxima at certain positions between the nuclei. In the vicinity of these maxima, the electron density exhibits all the properties normally associated with an atomic centre, except of course that there is no nucleus. These maxima have been termed non-nuclear attractors or even pseudo-atoms. In lithium clusters, the lithium atoms (nuclear attractors) appear to be bonded to these pseudo-atoms (non-nuclear attractors) rather than directly to one another, as one would normally

1. Boulton, G. & Clark, C. *Nature* **346**, 813–817 (1990).
2. Budd, W.F. & Smith, I.N. *Ann. Glaciol.* **3**, 42–49 (1982).
3. Oerlemans, J. *Nature* **297**, 550–553 (1982).
4. Huybrechts, P. & Oerlemans, J. *Ann. Glaciol.* **11**, 52–59 (1988).
5. Payne, A. & Sugden, D. *Ann. Glaciol.* **14**, 232–237 (1990).
6. Broccoli, A.J. & Manabe, S. *Géographie Physique et Quaternaire* **41**, 291–299 (1987).
7. Rind, D., Peteet, D. & Kukla, G. *J. geophys. Res.* **94**, 12851–12871 (1989).
8. Hays, J.D., Imbrie, J. & Shackleton, N.J. *Science* **194**, 1121–1132 (1976).
9. Neeman, B.U., Ohring, G. & Joseph, J.H. *J. geophys. Res.* **93**, 11153–11174 (1988).
10. Pollard, D., Muszynski, I., Schneider, S.H. & Thompson, S.L. *Ann. Glaciol.* **14**, 247–251 (1990).
11. Domack, E.W., Jull, A.J.T., Anderson, J.B., Linick, T.W. & Williams, C.R. *Quat. Res.* **31**, 277–287 (1989).
12. Hambrey, M.J., Ehrmann, W.U. & Larsen, B. in *Proc. Ocean Drilling Program. Sci. Results, Leg 119* (eds Barron, J. et al.) (in the press).
13. Jones, G.A. & Keigwin, L.D. *Nature* **336**, 56–59 (1988).

expect. Threefold coordination of these unusual sites appears to be particularly favourable.

For the smallest lithium cluster,  $\text{Li}_2$ , Cioslowski has described how the electron density can be characterized by a function very similar to the equation above illustrating the fold catastrophe, provided  $x$  represents the distance from the centre of the molecule and  $b$  is a quantity which depends only on the distance between the two Li nuclei. If the two atoms are very far apart,  $b$  is positive, and the pattern of critical points is typical of an ordinary molecule. As the atoms come closer together, a fold catastrophe is triggered at the critical distance for which  $b = 0$ . For shorter nuclear separations,  $b$  is negative and the pseudo-atom regime prevails, with minima placed symmetrically about a maximum at  $x = 0$ . The equilibrium geometry of this molecule corresponds to a bond length in this  $b < 0$  region, and so the nature of the bonding is entirely different from the picture usually presented. It is likely that this pseudo-atom picture will carry over to very large clusters of alkali metal atoms — that is, bulk metals — which are normally described by very different models.

Under ordinary conditions, methyl-lithium ( $\text{CH}_3\text{Li}$ ) usually exists as a tetramer or hexamer. The lithium atoms form a tetrahedron or octahedron with a

methyl ( $\text{CH}_3$ ) group above each triangular face, and it is tempting to imagine the interaction of an unpaired electron from each methyl group with a pseudo-atom on each triangular face. Small polyhedra of metal atoms bonded to various ligands (groups of atoms) occur widely for many transition metals, and these are being extensively studied experimentally by many research groups because of their potential use as catalysts for important industrial processes. It is plausible that the pseudo-atom picture might also be of relevance here and could contribute significantly to our understanding of the chemistry of such systems. All in all, a re-evaluation of current theoretical models for bulk metals and for systems based on clusters of metal atoms would appear to be very timely.  $\square$

David L. Cooper is in the Department of Chemistry, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK.

1. Thom, R. *Structural Stability and Morphogenesis* (Benjamin, New York, 1975).
2. Poston, T. & Stewart, I. *Catastrophe theory and its applications* (Pitman, London, 1978).
3. Cioslowski, J. *J. phys. Chem.* **94**, 5496–5498 (1990).
4. Gatti, C., Fantucci, P. & Pacchioni, G. *Theor. Chim. Acta* **72**, 413–458 (1978).
5. Cao, W.L., Gatti, C., MacDougall, P.J. & Bader, R.F.W. *Chem. Phys. Lett.* **141**, 380–385 (1987).
6. Cooper, D.L., Gerratt, J. & Raimondi, M. *Molec. Simul.* **4**, 293–312 (1990).

## TELOMERES

# All's well that ends well

Andrew Murray

SINCE the pioneering work of Muller<sup>1</sup> and McClintock<sup>2</sup> it has been known that telomeres, the ends of eukaryotic chromosomes, have special properties. Two papers on pages 866 and 868 of this issue<sup>3,4</sup>, together with a paper in *Nature* earlier this year<sup>5</sup>, shed light on what happens to the ends of chromosomes as cells age and how broken chromosomes can acquire new telomeres.

Telomeres exist to solve two problems: they must allow the complete replication of chromosome ends, but prevent the ends from fusing with each other or reacting with the internal regions of chromosomes. In all organisms that have been studied so far the ends of chromosomes consist of tandem repeats of short simple sequences, here depicted as  $T_nG_m$ , in which G and T residues predominate on the strand whose 3' end is at the telomere<sup>6</sup>. The number of repeats at each end is determined by the relative rates at which repeat sequences are gained and lost. Telomeric repeats are lost in each cell cycle because of the requirement to prime DNA synthesis with an RNA molecule that is ultimately discarded. Repeats are gained by the action of telomerase, an enzyme that can add telomeric repeats to the 3' end of a stretch

of telomeric repeats (see ref. 6 for a recent review) and, under some circumstances, by recombinational replication mechanisms that can copy repeats from the telomere of one chromosome to that of another<sup>7</sup>.

The balance between these forces has been extensively studied in unicellular eukaryotes. Some organisms such as *Tetrahymena*<sup>8</sup> and trypanosomes<sup>9</sup> adopt a 'binge and purge' strategy: both show a continuous increase in telomere length during vegetative growth, suggesting that the addition of telomeres exceeds their loss. Clearly, such a state of affairs cannot last for ever: telomere length must decrease eventually. In *Tetrahymena*, telomere length is reset during conjugation when new telomeres are formed by the action of telomerase<sup>8</sup>. Other organisms, such as the budding yeast, adopt a more moderate approach: telomere length fluctuates about a constant level during long-term culture, suggesting that addition and loss of repeats occur at equal rates<sup>10</sup>.

The identification of the vertebrate telomeric repeat<sup>11</sup>, TTAGGG, has allowed two groups to investigate the fluctuations in the length of human telomeres

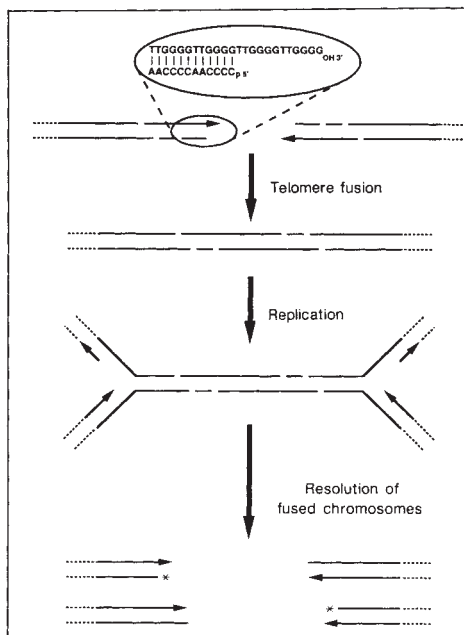
with age. Harley *et al.*<sup>5</sup> examined the effect of donor age and cell passage on telomere length in human fibroblasts, whereas Hastie *et al.*<sup>3</sup>, in their work reported in this issue, followed changes in telomere length in human tissues with age. Both groups report a surprising result: during the replication of somatic human cells, telomere length progressively decreases, demonstrating that the loss of telomeric repeats exceeds the rate of addition. In human tumours, which have undergone many more cell divisions than the tissue that gave rise to them, the telomeres are shorter than they are in matched normal tissue<sup>3,12</sup>. These results are all consistent with the suggestion that telomerase is inactive in somatic cells<sup>13</sup>.

To minimize the rate at which DNA is lost from the telomeres in the absence of telomerase, the  $T_nG_m$  sequences at the 3' ends of the chromosomes must be very efficient templates for DNA primase. This consideration raises the heretical possibility that the telomeric repeat sequences were initially determined by the sequence preferences of DNA primase, rather than by the special structural properties of G-rich sequences (reviewed in ref. 14). Intriguingly, bacteriophage T4 primase does use a  $T_nG_m$  sequence as its template<sup>15</sup>, and in *Oxytricha* a primase activity has been detected that can use the telomeric repeat as a template<sup>16</sup>. It will be interesting to see if the telomeric repeats are, in fact, preferred templates for the primases of eukaryotic cells.

The progressive loss of telomeric sequences during cell division may lead to chromosomal abnormalities. The *estI* mutation in budding yeast, for example, shows a gradual diminution in telomere length accompanied by a progressive increase in chromosome loss and mortality<sup>17</sup>. The loss of telomeric sequences can lead to end-to-end chromosome fusions by two routes: first, non-telomeric sequences can be exposed and then ligated to one another or to telomeric sequences, or telomeric sequences may become ligated to each other (telomere–telomere fusion). Many of the fused chromosomes will be dicentric and therefore participate in breakage–fusion–bridge cycles leading to the loss or duplication of genetic information and possibly increasing the rate of tumour progression<sup>3</sup>.

One structural feature of telomeres may have been designed to avoid telomere fusion. In *Tetrahymena*<sup>18</sup> and several other organisms, the telomeric repeats are interrupted by single-stranded gaps. If a fusion event involving a telomere occurs before the telomeric DNA has been replicated, replication past these gaps will separate the fused chromosomes (see figure). If such a strategy is used, the late replication of telomeres would help to maximize the chances of rescuing telomere fusions.





A scheme for the resolution of telomere-telomere fusions by replications through single-stranded gaps present in telomere repeat sequences. Telomeric sequences are shown as solid lines and internal chromosomal sequences as dashed lines. The inset shows the sequence organization of the end of the *Tetrahymena* telomere. Telomere-telomere fusion creates a dicentric chromosome, but this fusion is resolved when replication forks pass through regions of telomere repeat sequences that contain single-stranded gaps. No telomerase activity is represented, and therefore some sequence information is lost at the 5' end of the newly synthesized DNA strands (\*). Fusions of telomeres to non-telomeric ends are also resolved but in this case the non-telomeric end acquires a short stretch of inverted telomeric repeats.

Once the breakage-fusion-bridge cycle has begun, it perpetuates itself by continually generating new, reactive, non-telomeric DNA ends. This cycle can be broken if the new chromosomal ends become unreactive, a process referred to as healing. With the telomeric repeat sequences in hand, the molecular nature of healing events has now been investigated in the malarial parasite, *Plasmodium falciparum*<sup>19</sup>, and man<sup>4</sup>. In both cases healing has occurred by the acquisition of telomeric repeats at a site that previously lay deep within a chromosome.

The human healing event was discovered by Wilkie *et al.*<sup>4</sup> in a person with thalassaemia who carried a chromosome on which  $\alpha$ -globin genes were present but not expressed. Molecular analysis showed that this chromosome had acquired a new telomere just upstream of the  $\alpha$ -globin locus. There is an abrupt transition between the sequence at the globin locus and a stretch of human telomeric repeats. The stability of this chromosome over at least two generations shows that the telomeric repeat sequence is sufficient to form a functional chromosomal telomere even in the absence of more complex telomere-associated sequences, a conclusion that had

previously been reached with yeast<sup>20</sup>.

Does the sequence at which the new telomeric repeats have been added reveal anything about the nature of the healing event? Intriguingly, a G- and T-rich sequence within 10 base pairs of the site of telomere addition has a marked similarity to the human telomere repeat sequence. In yeast, telomeric repeats can be added to a non-telomeric sequence provided that telomeric repeats are within 30 base pairs of the end of the DNA molecule<sup>21</sup>, suggesting that telomerase may have a substrate recognition site spatially distinct from the site at which new telomeric repeats are added.

Finally, two studies in *Drosophila* suggest that not all broken chromosome ends are highly reactive<sup>22,23</sup>. In both cases, terminal deletions are maintained for many generations with a slow loss of DNA from the ends of the chromosomes. These ends lack any sequences that could be recognized as telomeric repeats, showing that under certain circumstances non-telomeric ends can be relatively inert. □

Andrew Murray is in the Department of Physiology, University of California, San Francisco, Box 0444, Parnassus Avenue, San Francisco, California 94143, USA.

## Fibrillated food

DIETARY fibre, that fashionable food component, has two virtues. First, it provides no calories, being indigestible. Second, it is a healthy tonic for the alimentary canal. Some foodstuffs high in amylose starch, like rice and spaghetti, actually increase in fibre content on being cooked. The amylose molecules aggregate into crystallites which defeat our digestive enzymes.

Daedalus is now generalizing this observation. He is devising food additives which subtly convert other components of our diet into indigestible fibre. DREADCO's chemists have set up shop in the firm's canteen and, over the protests of the cooks, are submitting selected meals to some simple chemical tricks. Traces of formaldehyde should cross-link the protein of a foodstuff into insoluble, enzyme-resistant polymers. Carbonyl chloride might couple its sugars into polycarbonates, while sulphur should usefully vulcanize its unsaturated fats. As additives, these crude and brutal reagents are unlikely to please the regulatory authorities. But in carefully controlled traces they will help Daedalus to discover how much of any given foodstuff can be converted to insoluble fibre without degrading its taste and texture too much. Since the taste of a food depends mainly on trace flavour substances, while its texture should be little affected by the microscopic polymer aggregates of new fibre, quite a lot of it may be 'fibrillated' without any detectable effect. Once the feasibility of the process has been demonstrated, subtler and more acceptable fibrillating additives can be developed for the job.

DREADCO's Fibrillated Foods should be widely welcomed by health freaks of all persuasions. Cream cakes, savoury pies, heavy puddings, all the richest and wickedest of cordon-bleu creations will soon be available in virtuous low-calorie, high-fibre versions. With sufficiently cunning chemistry, fast-acting fibrillating reagents could even be deployed as novel seasonings or condiments. Domestic cooks could outflank the family's bad eating habits by subtly fibrillating its meals, and guilty fast-food addicts could upgrade their hasty purchases to approved dietary standards.

But Daedalus senses a streak of masochism and puritanical fervour on the fringes of the health-food market. For the true vegetarian jogger, health is won by suffering. To capture this market, he is studying quite severe levels of dietary vulcanization. The DREADCO team is cross-linking the most delicate lettuce leaves into chewy tastelessness, reinforcing nuts and lentils to a sort of vegetable gravel, and devising a vegeburger which almost meets military specifications. Even the inherent wickedness of meat may be redeemed by brutal enough treatment.

David Jones

- Muller, H.J. *Collecting Net* **13**, 181-198 (1938).
- McClintock, B. *Proc. natn. Acad. Sci. U.S.A.* **25**, 405-416 (1939).
- Hastie, N.D., Dempster, M., Dunlop, M.G., Thompson, A.M., Green, D.K. & Allshire, R.C. *Nature* **346**, 866-868 (1990).
- Wilkie, A.O.M., Lamb, J., Harris, P.C., Finney, R.D. & Higgs, D.R. *Nature* **346**, 868-871 (1990).
- Harley, C.B., Fletcher, A.B. & Greider, C.W. *Nature* **345**, 458-460 (1990).
- Blackburn, E.H. *J. biol. Chem.* **265**, 5919-5921 (1990).
- Pluta, A.F. & Zakian, V.A. *Nature* **337**, 429-433 (1990).
- Larson, D.D., Spangler, E.A. & Blackburn, E.H. *Cell* **50**, 477-483 (1987).
- Bernards, A., Michels, P.A.M., Lincke, C.R. & Borst, P. *Nature* **303**, 592-597 (1983).
- Shampay, J. & Blackburn, E.H. *Proc. natn. Acad. Sci. U.S.A.* **85**, 534-538 (1988).
- Moysis, R.K. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**,

- 6622-6626 (1988).
- de Lange, T. *et al. Molec. cell. Biol.* **10**, 518-527 (1990).
- Cooke, H.J. & Smith, B.A. *Cold Spring Harb. Symp. quant. Biol.* **51**, 213-219 (1986).
- Frank-Kamenetski, M. *Nature* **342**, 737 (1989).
- Chai, T.A. & Alberts, B.M. *J. biol. Chem.* **261**, 7001-7010 (1986).
- Zahler, A.M. & Prescott, D.M. *Nucleic Acids Res.* **15**, 6299-6317 (1989).
- Lundblad, V. & Szostak, J.W. *Cell* **57**, 633-643 (1989).
- Blackburn, E.H. & Gall, J.G. *J. molec. Biol.* **120**, 33-53 (1978).
- Pologe, L.G. & Ravetch, J.V. *Cell* **55**, 869-874 (1988).
- Murray, A.W. & Szostak, J.W. *Molec. cell. Biol.* **6**, 3166-3172 (1986).
- Murray, A.W., Claus, T.E. & Szostak, J.W. *Molec. cell. Biol.* **8**, 4642-4650 (1988).
- Blessman, H., Carter, S.B. & Mason, J.M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1758-1761 (1990).
- Levis, R.W. *Cell* **58**, 791-801 (1989).

# Damage at a distance

SIR—The Loma Prieta earthquake registered 7.1 on the Richter scale. It occurred at 17:04 hours on 17 October 1989, along the San Andreas fault in the Santa Cruz mountains. Surprisingly, it destroyed about 40 houses in the Marina district of San Francisco, 67 miles away from the epicentre. Here, I would like to suggest an explanation for this damage.

In the 5 minutes after the main shock, many residents of the Marina district observed a mixture of sand, water and mud surfacing in their garages or backyards. In one apartment building 1 metre of sand oozed through the concrete floor. Many people reported that the mixture of water and sand emanated from the cracks and from the drainage openings of their garage floors. The sand ejecta, also called sand boils, sand spouts, water spouts, craterlets or sand volcanoes, were caused by the phenomenon of liquefaction, a physical process that transforms a solid mass of wet sands into a liquid mixture of sand and water<sup>1</sup>.

Figure 1 represents the 900 × 600 metre area of the Marina district which my students and I investigated. Located between the Presidio military reservation and Fort Mason, the area extends from Fillmore to Baker Streets from east to west, and from Marina Boulevard to Chestnut Street from north to south. We recorded a total of 74 sand boils in this area, the largest of which had a volume in excess of 3.5 cubic metres (Fig. 2). The ejecta were made of a grey and odourless mixture of fine dune sand and San Francisco bay mud.

The sand boils covered only the northeast corner of the Marina district (Fig. 1), a distribution related to the construction

history of the district. By 1894, a sea wall and a lagoon had been constructed along the original 1892 shoreline of the district. During the 1906 earthquake, Lawson reported<sup>2</sup> a severe ground shaking near this lagoon. By 1915, the lagoon had been filled to host the Panama-Pacific international exposition, after which some buildings were erected on the debris left over while others were directly built on man-made land. The Marina district was therefore constructed on a liquefiable landfill pre-disposed to ground motion amplification.

All our 74 observations on sand boils fall within the perimeter of the 1906 lagoon. The soils filling the lagoon were saturated by the high tide during the Loma Prieta earthquake. They liquefied, spread laterally and moved relative to the edge of the lagoon. As shown in Fig. 1, three apartment buildings collapsed and a serious fire was ignited along the shoreline of the old lagoon.

J. P. BARDET

*Department of Civil Engineering,  
University of California,  
Los Angeles,  
California 90089-0242, USA*

1. National Research Council. *Liquefaction of Soils during Earthquakes* 18–20; 72–73 (National Academy Press, 1985).

2. Lawson, A.C. et al. *Rep. State Earthquake Invest. Comm.* 404–405 (Carnegie Institution, Washington, DC, 1908).

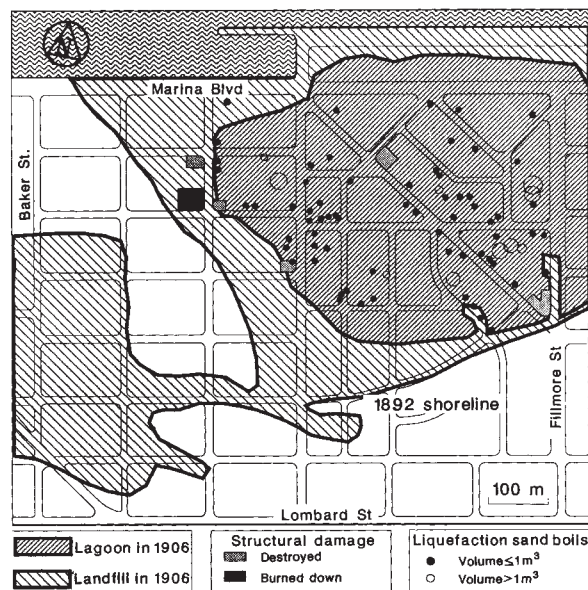


FIG. 1 The 1892 shoreline, the landfill area, the lagoon in 1906, the liquefaction sand boils and the collapsed buildings of the Marina district of San Francisco.

## Expanding zeolite-type nets

SIR—Zeolite materials have structures based on a framework of corner-connected tetrahedra with primarily silicon, aluminium or phosphorus at their centres and two-coordinated oxygen atoms at the corners. More than 60 topologically distinct frameworks of this type have been found so far in minerals and synthetic phases. Frameworks of this kind belong to the family of (4;2)-connected three-dimensional nets, of which an infinite number can be conceived<sup>1</sup>. I would like to draw attention to a beryllium-containing structural unit, which could in theory be used to 'expand' any zeolite-type net, resulting in a structure of low density and large pore size.

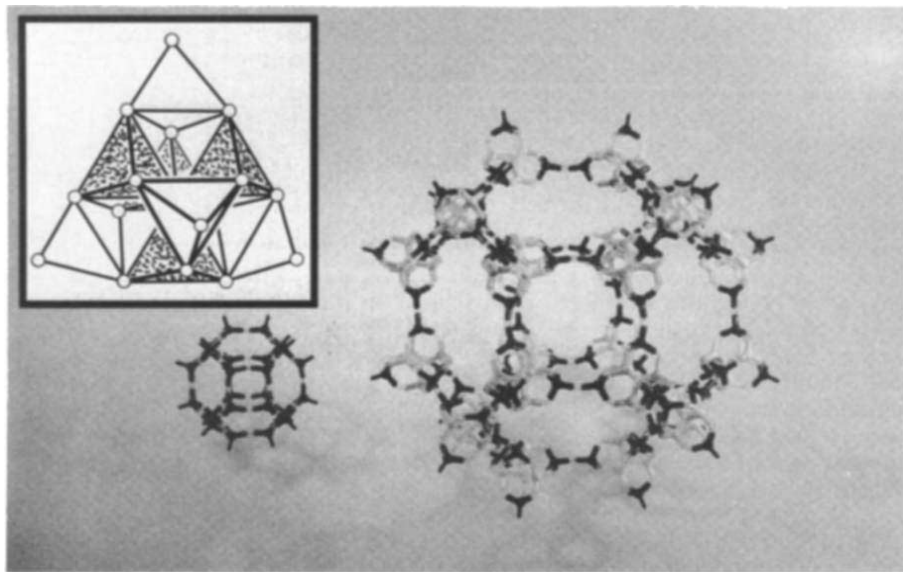
Zeolites possess pores of well defined dimensions that are of importance in catalysis and in separation processes. Among the aluminosilicates, the largest known apertures consist of 12 tetrahedra; low-density nets with larger rings would have great potential for industrial applications. Theoretical structure-building mechanisms that allow the expansion of a few specific nets have been presented<sup>2,4</sup> and one of these expanded nets, with 18-membered rings, has now been found<sup>5</sup> in the aluminophosphate VPI-5.

The isolated  $[\text{Be}_4\text{Si}_4\text{O}_{17}]^{10-}$  ion<sup>6</sup> exists in the compound  $\text{Na}_{10}\text{Be}_4\text{Si}_4\text{O}_{17}$  (inset in the figure). In the centre of the ion, four  $\text{BeO}_4$  tetrahedra share an oxygen atom. The other three oxygen atoms of each  $\text{BeO}_4$  group are two-coordinated through corner-sharing between  $\text{BeO}_4$  and  $\text{SiO}_4$  tetrahedra. Three of the corners of a  $\text{SiO}_4$  tetrahedron are shared with  $\text{BeO}_4$  tetra-



FIG. 2 A large sand boil (3.5 cubic metres) with two craters emerged in a backyard a few minutes after the main shock. The mixture of water, sand and clay flowed through the wooden fence and completely covered the neighbouring backyard.





Inset is sketch of  $[\text{Be}_4\text{Si}_4\text{O}_{17}]^{10-}$  ion (ref. 6):  $\text{BeO}_4$  tetrahedra are shaded and  $\text{SiO}_4$  tetrahedra unshaded. Main photo shows the cage in the sodalite structure (left), and in expanded form (right).

hedra, and the fourth corner is free. This polyanion has the same symmetry as a regular tetrahedron (point group  $T_d$ ) and can be regarded as a 'supertetrahedron' comprised of eight tetrahedra.

If every tetrahedron in a zeolite net is replaced by this supertetrahedron, a new, expanded net is generated. This procedure can be repeated an infinite number of times, but only the first expansion will be considered here. If the unit cell of the original net contains  $n$  tetrahedra, the expanded net will contain  $8n$  tetrahedra and the unit-cell content will change from  $\text{T}_n\text{O}_{2n}$  to  $\text{T}_{8n}\text{O}_{16n}$ . The maximum symmetry of the net will not change, but the multiplicity of the rings will increase by a factor of three: for example, the 4- and 6-rings in the sodalite-type frame are transformed into 12- and 18-rings (see figure). The framework density (number of tetrahedra per  $\text{nm}^3$ ) ranges from 12.5 to 29.0 in known framework structures<sup>7</sup>. Introduction of the supertetrahedron involves a linear expansion by a factor of about 2.7, and thus the original framework density is multiplied by a factor of  $8 \times 2.7^{-3} = 0.41$  (assuming a rigid framework). Each supertetrahedron contains no less than twelve 3-rings, in agreement with the conclusion<sup>7</sup> that very open tetrahedral networks (with two-coordinated oxygen atoms) contain a maximum number of 4- and/or 3-rings. The four-coordinated oxygen atom in the  $\text{T}_4\text{O}_{17}$  unit would bring a novel feature to the field of zeolite-type nets.

STAFFAN HANSEN

*Inorganic Chemistry 2, Chemical Center,  
University of Lund,  
PO Box 124,  
S-221 00 Lund, Sweden*

1. Smith, J.V. *Chem. Rev.* **88**, 149–182 (1988).
2. Barrer, R.M. & Villiger, H. Z. *Kristallogr.* **128**, 352–370 (1969).

3. Smith, J.V. & Dytrych, W.J. *Nature* **309**, 607–608 (1984).
4. Meier, W.M. *Pure appl. Chem.* **58**, 1323–1328 (1986).
5. Davis, M.E. *et al. Nature* **331**, 698–699 (1988).
6. Frostäng, S. *Chem. Comm. Univ. Stockholm No. 3* Jan. 29, 1–72 (1990).
7. Brunner, G.O. & Meier, W.M. *Nature* **337**, 146–147 (1989).

## Cosmology theory compromised

SIR—A curious dynamic tension has been rising in the field of cosmology. Some widely held theoretical assumptions are coming into increasing conflict with observational results, and yet those assumptions continue to receive strong support. Consider three prime examples.

(1) Various empirical tests suggest that the cosmological mass density parameter  $\Omega$  is less than 1. Theoreticians have argued that  $\Omega$  almost certainly has a value of 1. This conclusion stems from the fact that the 'inflationary scenario', which rescues the Big Bang theory from several serious difficulties, and the cold dark matter theory, which is used to explain the formation of galactic scale structure, are highly dependent upon the value of 1 for  $\Omega$ . But, an objective weighing of the empirical evidence<sup>1,2</sup> on  $\Omega$  suggests that  $\Omega \leq 0.3$ . Do we put our trust in theoretical assumptions or in observational results?

(2) Recent observations seem to favour baryonic dark matter. Hypothetical non-baryonic particles, such as weakly interacting massive particles (WIMPs), have been vigorously advocated as the most likely candidate for the unseen dark matter which dominates the mass of the Universe. But there have been many non-detections and ever-narrowing empirical constraints. A remarkable set of observations by Tyson *et al.*<sup>3</sup> show that the distribution of the dark matter in tested galactic clusters closely parallels that of

cluster red light (baryonic matter). The most straightforward interpretation of this result is that the dark matter is baryonic. Given that there is no empirical evidence for non-baryonic dark matter, and some evidence against it<sup>1,3</sup>, should it still be our best candidate?

(3) Something seems to be amiss with the Big Bang paradigm. This paradigm has always faced some theoretical problems: the horizon problem, the flatness problem, the smoothness problem and a postulated acausal beginning to the Universe. But a new set of empirical challenges has recently arisen: a Universe dominated by unpredicted dark matter, unanticipated large-scale galactic streaming, globular star clusters that are estimated to be as old as  $18 \times 10^9$  years, large scale ( $\geq 10^2$  megaparsec) structure<sup>4</sup> far exceeding original predictions and, ironically, a microwave background radiation that is too homogeneous.

The inflationary model and the cold dark matter model were created to save the Big Bang paradigm from conflict with many of these findings. But given the difficulties mentioned above with which these supplementary theories are now beset, their ability to aid the Big Bang paradigm is in some doubt.

At this crucial juncture comes an observational result<sup>4</sup> that leaves almost everyone scratching their heads: the apparent detection of a periodic pattern of galactic clustering that is maintained over scales of the order of  $10^3$  megaparsecs! Certainly, it appears that the observable part of the Universe has expanded from a far more compact state, beginning about  $10^{10}$  years ago. But beyond these very modest empirical constraints lies speculation and the unknown. Do we continue to pretend that we have just about got the Universe figured out?

The underlying theme of these three examples is the idea that we may have already passed the point at which it is appropriate for observational results to take precedence over theoretical constructions. There is not yet enough empirical evidence for conclusive scientific decisions on these three issues, but surely science would be better served if a greater attempt was made to loosen the grip of prevailing prejudices.

ROBERT L. OLDERSHAW

Box 2262,  
Amherst College, Amherst,  
Massachusetts 01002, USA

1. Geller, M. & Huchra, J. *Science* **246**, 897–903 (1989).
2. Dressler, A. in *Astronomy, Cosmology, and Fundamental Physics* (eds Caffo, M., Fanti, R., Giacomelli, G. & Renzini, A.) 23–40 (Kluwer, Dordrecht, 1989).
3. Tyson, J.A., Valdes, F. & Wenk, R.A. *Astrophys. J.* **349**, L1–L4 (1990).
4. Broadhurst, T.J., Ellis, R.S., Koo, D.C. & Szalay, A.S. *Nature* **343**, 726–728 (1990).

■ See also 'The extragalactic Universe: an alternative view', page 807.

## Streptomycin and self-splicing

SIR—The aminoglycoside antibiotic streptomycin, which interferes with prokaryotic protein synthesis by interacting with the ribosomal RNA, is also able to interact with the RNA of an autocatalytic group I intron. It is likely to do so by competing with the substrate guanosine via the guanidino group, which guanosine and streptomycin have in common.

Group I introns, like the *Tetrahymena thermophila* rRNA and the T4 phage *td* introns, have the ability to undergo self-splicing *in vitro* by a two-step reaction<sup>1,2</sup>. The first step requires binding of the substrate guanosine, which becomes covalently bound to the first nucleotide of the intron after cleavage of the 5' splice site, and the second involves cleavage at the 3' splice site and ligation of the exons. The cofactor guanosine binds to the ribozyme by hydrogen bonds using the guanidino group<sup>3</sup>. This guanosine-binding site has been partially assigned to the conserved G-C base pair in the P7 helix (G264-C311 in the *Tetrahymena* intron)<sup>4</sup>.

Streptomycin contains two guanidino groups that could interfere with guanosine binding. We performed splicing reactions *in vitro* with increasing amounts of streptomycin and found that it can completely inhibit the splicing reaction at a concentration of 10 mM. Half-maximal inhibition was achieved at 5 mM, with guanosine at

its  $K_m$  concentration. Increasing the guanosine concentration restored splicing, suggesting competitive inhibition. The same results were obtained with the *Tetrahymena* rRNA intron. Similar competition with guanosine has been described for arginine, which also contains a guanidino group<sup>5</sup>.

Streptomycin directly interacts with the 16S rRNA of *Escherichia coli*<sup>6</sup> and interferes with translation initiation and proofreading in prokaryotic protein synthesis<sup>7</sup>. Both translation initiation and proofreading involve GTP binding and hydrolysis; it is not known if streptomycin interferes with GTP binding during these processes. Elucidating the mechanism of splicing inhibition by streptomycin may reveal how this antibiotic acts on the prokaryotic ribosome.

UWE VON AHSEN  
RENEE SCHROEDER

Institut für Mikrobiologie und Genetik,  
Universität Wien,  
Althanstrasse 14,  
1090 Vienna, Austria

1. Cech, T. R. *Science* **236**, 1532–1539 (1987).
2. Ehrenman, K. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5875–5879 (1986).
3. Bass, B. L. & Cech, T. R. *Biochemistry* **25**, 4473–4477 (1986).
4. Michel, F., Hanna, M., Green, R., Bartel, D. P. & Szostak, J. W. *Nature* **342**, 391–395 (1989).
5. Yarus, M. *Science* **240**, 1750–1758 (1988).
6. Moazed, D. & Noller, H. *Nature* **327**, 389–394 (1987).
7. Thompson, R. C. *Trends biochem. Sci.* **13**, 91–93 (1988).

## Kaposi's sarcoma and AIDS

SIR—The role of human immunodeficiency virus type 1 (HIV-1) in the pathogenesis of Kaposi's sarcoma is thought to be indirect. HIV-1 sequences are absent in DNA from cultured Kaposi's sarcoma-derived cells<sup>1</sup>, and transgenic mice carrying the *tat* gene and developing Kaposi's-like lesions express *tat* in their skin but not in the tumour cells<sup>2</sup>. In a recent Letter<sup>3</sup>, Ensoli et al. reported that the *tat* protein of HIV-1 is released extracellularly from HIV-1-infected cells *in vitro* and stimulates the growth of cells derived from Kaposi's lesions of AIDS patients<sup>3</sup>. The same growth-promoting properties were observed with recombinant *tat*, and were inhibited by anti-*tat* antibodies, independent of the source of *tat*<sup>3</sup>. These findings indicate that *tat* released from HIV-1-infected cells may contribute to the induction of Kaposi's in HIV-1-infected individuals. We therefore compared the prevalence of anti-*tat* antibodies in patients with AIDS whose AIDS-defining diagnosis did or did not include Kaposi's.

We assayed sera obtained and stored from 297 HIV-1 seropositive patients at the time of or within one month of

diagnosis of AIDS (according to Centers for Disease Control criteria) for anti-*tat* antibodies with a previously described enzyme immunoassay using *Escherichia coli*-produced *tat* as antigen<sup>4</sup>. The group with Kaposi's consisted of patients with a sole diagnosis of Kaposi's ( $n = 67$ ), as well as of patients in whom AIDS was diagnosed because of the concurrent presence of Kaposi's and an AIDS-defining opportunistic infection ( $n = 11$ ). The group without Kaposi's consisted of people presenting with an AIDS-defining opportunistic infection ( $n = 203$ , including two patients with concurrent non-Hodgkin's lymphoma), but also included patients who were diagnosed with AIDS solely on the basis of a non-Hodgkin's lymphoma ( $n = 12$ ) or AIDS-dementia complex ( $n = 4$ ).

### TAT-SPECIFIC ANTIBODIES IN AIDS

| AIDS-defining condition   | Kaposi's sarcoma    |           |
|---------------------------|---------------------|-----------|
|                           | present             | absent    |
|                           | No. of patients (%) |           |
| Anti- <i>tat</i> positive | 10 (13)             | 21 (10)   |
| Anti- <i>tat</i> negative | 68 (87)             | 198 (90)  |
| Total                     | 78 (100)            | 219 (100) |

The *tat*-specific antibody response was determined in serum obtained within one month of diagnosis of AIDS

In both groups, the prevalence of anti-*tat* antibodies was low (13 versus 10 per cent), without significant difference between them (Fisher's exact test  $P = 0.27$ ; see table). Evidently, HIV-1 infection in Kaposi's patients does not increase the antigenicity of *tat* relative to the infection in non-Kaposi's patients, pleading against overexpression of *tat* in Kaposi's patients. Also, the presence of anti-*tat* antibodies in 10 of the 78 Kaposi's patients clearly did not protect them from developing Kaposi's. Our data indicate that no evidence can be put forward for either high antigenicity of *tat* or a protective effect of anti-*tat* antibodies in patients with AIDS-related Kaposi's sarcoma.

PETER REISS  
JOEP M. A. LANGE

Department of Medicine, AIDS-unit,  
Academic Medical Centre,  
Meibergdreef 9,  
1105 AZ Amsterdam, The Netherlands

CARLA L. KUIKEN  
JAAP GOUDSMIT

Human Retrovirus Laboratory,  
Academic Medical Centre,  
Meibergdreef 15,  
1105 AZ Amsterdam, The Netherlands

ENSOLI AND GALLO REPLY—Reiss et al. describe data on anti-*tat* antibodies in HIV-1-infected individuals with or without Kaposi's sarcoma which, in their opinion, are sufficient to indicate "that no evidence can be put forward for either high antigenicity of *tat* or a protective effect of anti-*tat* antibodies in patients with AIDS-related Kaposi's." These conclusions are based on the detection of anti-*tat* antibodies in 13 per cent of Kaposi's patients with AIDS versus 10 per cent of AIDS patients without Kaposi's. In our opinion, these data are not sufficient to support any conclusions about the role of *tat* in Kaposi's for the following reasons.

(1) Antibodies directed against HIV-1 regulatory proteins may change in titre and disappear during HIV-1 infection<sup>5</sup>. Therefore, longitudinal studies are necessary for evaluating their significance.

(2) Antibodies against HIV-1 regulatory proteins are relatively weak compared with other HIV-1 proteins, indicating that these proteins may be poor antigens *in vivo*<sup>6</sup> or that they are present in low amounts.

(3) Reiss et al. do not show whether anti-*tat* antibodies detected in AIDS patients or in earlier stages of the disease, are neutralizing the biological activities of *tat*<sup>3,6</sup>. Therefore, the conclusion that they cannot prevent the occurrence of Kaposi's in HIV-1 infected individuals is still open.

(4) Reiss et al. do not show data concerning the presence of *tat* in the sera of the same patients. The two phenomena (protein and antibodies to them) may not be parallel and both may depend on the stage of the disease, the rate of infection



and the function of the immune system. They may also be dependent on antigen stimulation and T-cell activation, which may be more frequent in homosexuals in whom Kaposi's is more common.

(5) The presence or absence of anti-*tat* antibodies or *tat* protein in the sera of HIV-1-infected individuals obviously does not exclude the role of *tat* in the pathogenesis of Kaposi's because it could be released and taken up by target cells in close proximity or by cell-cell contact.

(6) Finally, if high levels of anti-*tat* antibodies were found, Reiss *et al.* could argue that high levels should be protective by inactivating *tat* as might occur, for example, in polio. Therefore, *tat* would have no role in Kaposi's. But antibodies are not always neutralizing, and the meaning of high levels can be just the opposite — high levels of antibodies to HTLV-I are not protective (unless present before infection). High levels, in fact, suggest the presence of more virus. Conversely, what do low antibodies mean? Would they indicate that not much *tat* is present; or that

there are insufficient neutralizing antibodies? Therefore, it is not clear to us what an average level of antibodies really means.

Thus, in our view, the titres cannot be used to draw any meaningful conclusions regarding the role of *tat* or the anti-*tat* antibodies *in vivo* in the pathogenesis of Kaposi's.

BARBARA ENSOLI  
ROBERT C. GALLO

Laboratory of Tumor Cell Biology,  
National Cancer Institute,  
National Institutes of Health,  
Bethesda, Maryland 20892, USA

1. Salahuddin, S.Z., Nakamura, S., Biberfeld, P., Kaplan, M.H., Markham, P.D., Larsson, L. & Gallo, R.C. *Science* **242**, 430–433 (1988).
2. Vogel, J., Hinrichs, S.H., Reynolds, R.K., Luciw, P.A. & Jay, G. *Nature* **335**, 606–611 (1988).
3. Ensoli, B., Barillari, G., Salahuddin, S.Z., Gallo, R.C. & Wong-Staal, F. *Nature* **345**, 84–86 (1990).
4. Goudsmit, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 4478–4482 (1988).
5. Devash, Y., Reagan, K., Wood, D., Turner, J., Parrington, M. & Kang, C.Y. *Nature* **345**, 581 (1990).
6. Viscidi, R.P., Mayur, K., Lederman, H.M. & Frankel, A.D. *Science* **246**, 1606 (1989).
7. Ensoli, B., Barillari, G., Salahuddin, S.Z., Gallo, R.C. & Wong-Staal, F. *Nature* **345**, 84 (1990).

## Sunlight brighter than the Sun

SIR—We have concentrated terrestrial sunlight by a factor of 84,000 to intensities of  $72 \text{ W mm}^{-2}$ , establishing a new world record for solar flux concentration that exceeds our previous result by a factor of 1.5 (refs 1,2). This concentration exceeds the intensity of light at the surface of the Sun itself, which is  $63 \text{ W mm}^{-2}$ ; we have thus produced the highest concentration of sunlight in the Solar System.

We use a device designed according to the principles of nonimaging optics<sup>3</sup>. A nonimaging concentrator is essentially a 'funnel' for light, and can concentrate light to intensities four times greater than those obtained with conventional imaging devices. (Aberrations in image-forming devices limit their ability to concentrate light.) A laser crystal or a solar furnace does not need to receive a perfect image of the Sun, rather the maximum power per area. By dispensing with image-forming requirements, we can attain concentrations that approach the theoretical limit of:

$$C_{\text{max}} = n^2/\sin^2\theta \quad (1)$$

where  $n$  is the index of refraction at the target surface and  $\theta$  is the semiangle subtended by the Sun. The limit can be derived from both phase space conservation<sup>4</sup> and thermodynamic arguments<sup>5</sup>.

Our concentrator is designed according to the 'edge-ray' method<sup>6</sup>, in which all light rays entering the device at the maximum collection angle are directed after one reflection at most to the rim of the exit aperture (see figure). The other rays are therefore reflected within the aperture itself.

A nonimaging concentrator alone,

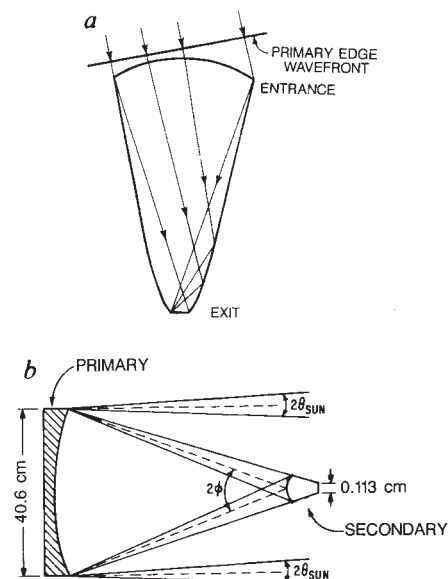
could attain high concentrations of sunlight, but would be large and unwieldy. We therefore use a two-stage system incorporating a parabolic mirror and a nonimaging concentrator. The overall concentration is the product of the concentrations of the parabolic mirror and of the nonimaging concentrator:

$$C = n^2 \cos^2\phi / \sin^2\theta \quad (2)$$

which falls short of the theoretical limit in equation (1) only by the factor of  $\cos^2\phi$ , where  $\phi$  is the rim angle of the mirror.

Using such a technique, we previously achieved a concentration of sunlight of 56,000 times the intensity at the surface of the Earth<sup>1</sup>. Our modified apparatus reaches still higher concentrations. We use a 40.6-cm-diameter silver-coated telescope mirror with rim angle  $11.5^\circ$  (focal ratio 2.5). Sunlight is concentrated to a 1-cm spot 1 m away from the mirror. The light is further compressed to a 1-mm spot by a sapphire nonimaging concentrator, chosen for its low absorption and high index of refraction,  $n = 1.76$ . Substituting these values into equation (2) gives the theoretical limit to the concentration of our new two-stage system, 137,000, which is a 31% increase over the theoretical upper limit of 104,000 in our earlier device.

Another modification is that our concentrator now works by total internal reflection. In the previous experiment the walls of the concentrator were silvered, and allowed losses due to absorption. Total internal reflection is, however, nearly loss-free, and increases the efficiency of the concentrator.



In a nonimaging concentrator designed by the edge-ray method (shown in *a*) all light rays entering the device at the maximum collection angle are directed after one reflection at most to the rim of the exit aperture. The edge-ray principle is slightly modified for the concentrator described here, to ensure that light is reflected by total internal reflection near the top. Hence, the exit angle of the light is reduced slightly, from  $90^\circ$  to  $86^\circ$ . Panel *b* shows the two-stage approach to concentrating sunlight.

In a series of trials over several cold, crisp days in January and February we measured an average power per area at the exit aperture of the sapphire concentrator of  $72 \text{ W mm}^{-2}$ . This intensity is 84,000 times greater than the direct incident intensity,  $0.86 \text{ mW mm}^{-2}$ .

Potential uses for such high levels of solar flux are only now being explored. In the past it has been difficult to make efficient solar pumped lasers because of the limited materials available and the lower levels of concentration of sunlight that could be achieved. We believe our approach should increase the attained efficiencies, with the aim of making a tunable solar-pumped laser. Other applications include the destruction of hazardous waste and high-temperature processing of specialized materials.

DAVE COOKE  
PHIL GLECKMAN  
HELMUT KREBS  
JOE O'GALLAGHER  
DAN SAGIE  
ROLAND WINSTON

Department of Physics and  
The Enrico Fermi Institute,  
The University of Chicago,  
Chicago, Illinois 60637, USA

1. Gleckman, P. *Appl. Opt.* **27**, 4385–4391 (1988).
2. Gleckman, P., O'Gallagher, J. & Winston, R. *Nature* **339**, 198–200 (1989).
3. Welford, W.T. & Winston, R. *High Collection Nonimaging Optics* (Academic, New York, 1989).
4. Winston, R. *J. opt. Soc. Am.* **60**, 245 (1970).
5. Rabl, A. *Sol. Energy* **18**, 93 (1976).

# Ghost in the machine

Struther Arnott

**Killing the Spirit. Higher Education in America.** By Page Smith. Viking: 1990. Pp. 315. \$19.95.

TESTILY, I sometimes remind patronizing marketmen that I am the chief executive officer of an adaptable enterprise that has been trading continuously for 580 years, and that therefore I must take a longer view than they of the products I should market. The better-informed or more argumentative of them — usually our own alumni — suggest that the St Andrews' market niche may be somewhat special and that the market as a whole may have been changed qualitatively by the explosive growth of university business in the past 50 years. In the autumn of 1939 for example, before the United States had committed itself to participating in the war, 7,119 students were enrolled by Purdue University, Indiana, to be taught by 809 faculty members, deploying a total budget of \$2,701,364. This year, the corresponding numbers for the West Lafayette campus were 35,817 students and 2,192 faculty members, with a \$651,443,740 budget.

The expansion of the academic supermarkets in the United States has been matched in the United Kingdom by growth in boutiques like St Andrews. In 1939, St Andrews, for all its considerable international reputation, was a lilliputian university with 536 students, 67 teachers and a budget of £114,328. Fifty years on, these numbers had grown to 4,184 students and 344 teachers, consuming £30,305,899. A moment's work with a calculator reveals a remarkable concordance in the 50-year inflation rates of

these apparently very different institutions. Clearly, a love affair with higher education has not been an exclusively US phenomenon.

Who can be surprised that academic obesity has attendant ills? Bureaucratization is both pervasive and assertive. In the largest universities in the United States it is now rare to find distinguished scholars occupying any leadership position, even the quasi-sacerdotal deanship of the graduate school. In the United Kingdom, universities have increasingly powerful bureaucratic minders, the secretary-registrars, who have their own national conference at which they recently felt compelled to offer advice directly to the British government explicitly on the grounds that the subject of interest was too urgent and too important to await the considered opinions of their principals/vice-chancellors/presidents or other colleagues in the universities.

Education and training based on intimate master-disciple or craftsman-apprentice relationships survive in diluted form at best in expanded universities. A significant demonstration of this weakened bond occurred in the 1988–89 national pay dispute in the United Kingdom when the teachers at only one university in the country were willing to put traditional academic commitments to their own students before solidarity with fellow trade unionists at other institutions.

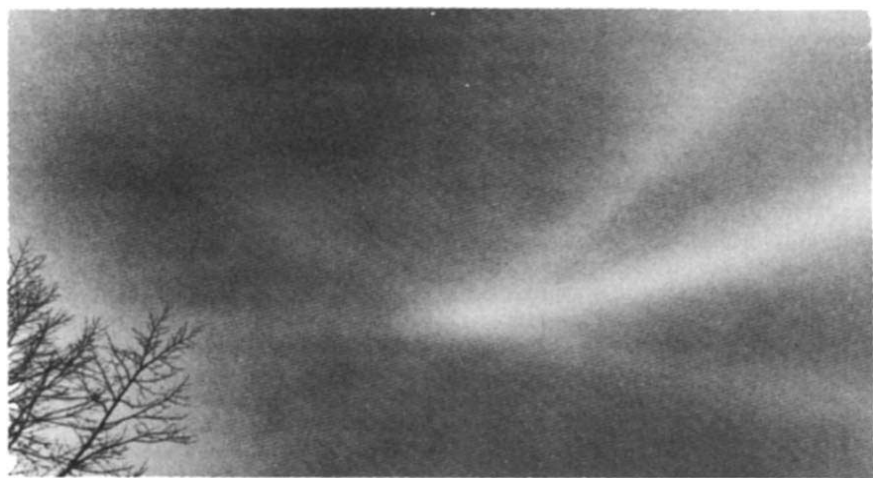
Intellectual corruption in research is

now not uncommon in the poorly managed academic latifundia supported by the US National Institutes of Health, and has been a lively issue in the US Congress. Even more worrying, I would have thought, is the meretriciousness of a great deal of non-fraudulent scholarly effort which, measured by citation interest, in too many cases has a half-life of less than a year. Perhaps not yet corrupt, but certainly corrupting, has been the ritualization and dilution of discrimination in many processes connected with staff selection, retention, reward and promotion.

It is of course easy to sneer at academics. They are uncommonly opinionated, frequently wordy, enjoy self-caricature and often perversely aspire to mimic their distorted inventions. Thereby they give to fortune legions of hostages that can be decimated by knowing antagonists in a few hundred pages of profitable prose. Other large, fast-growing businesses are more reticent about their similar problems and can be analysed only with great effort, given their comparative lack of transparency.

A practitioner like myself was bound to be intrigued by a book about US universities and their current predicaments, written from the perspective of a popular historian, whose first publication was about James Wilson, one of the St Andrews' alumni who were founding fathers of the United States, and whose latest publication is larded with many wise saws and modern instances of academic practices and spiced with first-hand experiences at institutions as diverse as Dartmouth, Harvard and the University of California at Los Angeles and at Santa Cruz. The book is written most engagingly, as one might expect from its highly professional and aptly named author Page Smith, who has to his credit nine volumes of (certainly not marxist) people's histories of the United States, eight book-of-the-month club main selections and one an alternate. He would have us believe that he "inadvertently" won a Rufus Choate scholarship in his senior year at Dartmouth and that he contrived to get into the history graduate programme at Harvard to evade the English programme's "stringent and irrelevant" requirements of knowledge of Greek, Latin, middle high German and Icelandic. We are, it seems, supposed to be impressed that Smith did not confirm these feared requirements at the time (nor, apparently, since), and having failed to strain these gnats we should be ready to swallow all the camels of selective quotation that he proceeds relentlessly to herd across more than 300 subsequent pages.

A book with the title *Killing the Spirit* might be expected to be a polemic, but in the end I was left with the impression that for all its apparent scholarship it is merely a prolonged sneer. And, being



One of the few photographs that has been obtained of an anhelical arc. The anhelical point lies 180° in azimuth away from the Sun but has the same elevation above the horizon as the Sun. This is one of many optical effects in the sky described in *Rainbows, Halos, and Glories* by Robert Greenler, which includes many colour illustrations. Publisher is Cambridge University Press, price is £12.95, \$22.95 (pbk). □



presented as a history it could, wittingly or unwittingly, evade any obligation to prescribe remedies for the maladies it chronicled. There is indeed a final chapter entitled "Reviving the Spirit", but in it continued diagnosis leaves little room for remedies. By this time one is not surprised, because there is another intriguing omission from Smith's history, which is any account of his efforts at reform when he held responsibility as founding provost at the University of California, Santa Cruz.

Is this book more than a sour farewell from a resentful or disappointed colleague? Perhaps not: even the most successful scholars in universities can harbour resentments over long periods of time. I can remember vividly the incredulity with which I watched and listened to the normally avuncular W. L. Bragg pointedly intercalating into the festive television programme in honour of the 50th anniversary of his Nobel prize (shared with his father W. H. Bragg) the fact that Bragg's law was not Braggs' law and that the singular Bragg in this case was W. L. and not W. H., who had insensitively exploited his seniority to be more visible at scientific meetings before World War I than his youthful collaborator and relative.

Even a self-indulgent book is written for readers. Who might they be in this case? Smith is not enthusiastic about the process by which tenure is earned but does not care to address the central issue of whether tenure is merely life-long job security for the incompetent or brass-bound protection for workers with unpopular opinions. He enjoys "publish or perish" as a slogan but not as a doctrine. (Who does?) But how to address the problem of auditing the productivity of the silent public pensioner is an issue he does not tackle. Smith resents the increasing weight of science and scientists and the lack of appropriate attention to the older "big questions". Do they, I wonder, require little Latin, less Greek and no middle high German or Icelandic? No molecular genetics, no biochemical anthropology, nothing of the wondrous discoveries that might have frightened but also intrigued Smith as they occurred in his own lifetime?

I have to conclude that this book is designed only for academic moths with unrealizable desires for the lights of mythical stars. It might bring comfort to those for whom David Lodge's fictional satires of the academy are too frivolous, but it will not help anyone concerned to reform universities or dedicated to forging new treaties between those who make and implement public policies for higher education and training. □

*Struther Arnott, Principal and Vice-Chancellor of St Andrews University, St Andrews, Fife KY16 9AJ, UK, has worked and still works in universities on both sides of the Atlantic.*

## Never apologize. . .

Chris Paul

**Evolutionary Paleobiology of Behavior and Coevolution.** By A. J. Boucot. *Elsevier*: 1990. Pp. 725. \$148.75, £90.62.

PALAEONTOLOGY suffers from several misperceptions which, because they are often propounded by palaeontologists, make us our own worst enemies. If we don't believe in our subject, who else will? First, palaeontology is thought to consist of little more than documenting fossils, describing their parts and assigning ages to the rocks that contain them. Then there is the legendary incompleteness of the fossil record which can be, and is, used as an excuse to reject any fossil evidence that runs counter to current prejudices. Occam's razor alone should lead us to accept the simplest hypothesis consistent with the facts — that the fossil record faithfully, if incompletely, preserves what actually happened. Yet practitioners of this art, like the beneficiaries of the Official Secrets Act, find it too convenient to abandon. Finally, palaeontology is said to offer less information than neontology (the study of living organisms). To some palaeontologists, fossils preserve information so different from that derivable from living organisms that the two cannot be classified in the same way. Boucot is not an apologist for palaeontology, yet his book contains its own paradoxes. In the preface he states that one aim was to counter the first misperception and in the introduction we learn that he is well aware of the second, yet the book most effectively counters the third. This is its greatest strength.

In 1977, Boucot began collecting material for a journal article. Since then, like Topsy's, his collection has grown into a book of more than 700 pages — and therein lies a significant weakness. The contents are largely a review of the literature supplemented by additional comments from many other contributors. The bulk of the book, "paleontologic evidence", is presented under 35 separate headings. Treatment, choice and order of headings seem rather arbitrary. For example, it is reasonable to devote a large section to "predation and feeding behaviors" (*sic*; 143 pages and by far the largest section), but why only two pages on functional morphology? Why have a section on "specialized, potentially interacting biologic substrates" separated by six sections from another entitled "specialized substrates"? Sections vary from the very generalized, "trace fossils and their formers", to the highly specific, "anuran chromatophores", "fighting birds" or "lateral line pressure receptors". Fourteen of the thirty-five sections occupy one page

or less and the reader loses track of a coherent theme. In places, the book is in danger of becoming a "wonder book of fascinating fossil facts". Yet it is much more than that. One special feature is the definition in the introduction of six categories of reliability for the interpretations of behaviour, ranging from incontrovertible "frozen behaviour", such as mating insects trapped in amber, to the highly speculative. Even the latter are justified. Speculation on the antiquity of glochidia (parasitic larval stages of unionacean freshwater bivalves) seems reasonable even if fossil glochidia are confined to the Pleistocene. Adult unionaceans are known from the Mesozoic and such speculation should stimulate the search for older fossil glochidia.

The concluding sections return to the main theme: the now classic debate on rates and modes of evolution, and present the book's chief paradox. Although Boucot concludes that community evolution is effectively punctuational — new "community groups" arise in 1–2 million years and persist for tens of millions of years — there is decided scepticism about punctuated equilibrium itself. In an appendix, Koch reconsiders Eldredge's original trilobite data, but states "The absence of positive examples favoring the punctuated equilibrium concept is a major factor militating against its acceptance"! Another problem is the danger of circular arguments. Boucot believes behaviour is closely correlated with taxonomy at the familial level, yet this is almost inevitable. Radically new behaviour will be accompanied by significantly different morphology and will, in turn, precipitate the recognition of new families. One wonders if Boucot's "community groups" aren't just families that appear suddenly and last for tens of millions of years, which brings us back to punctuated equilibria . . . Nevertheless, the book is crammed full of examples of palaeobiology at its best and shows conclusively that palaeontology need not be a second-class life science if only we palaeontologists could stop apologizing for it. □

*Chris Paul is in the Department of Earth Sciences, Liverpool University, Brownlow St, Liverpool L69 3BX, UK.*

■ Two new books address the issues of evolution and creationism. One is *Evolution and the Myth of Creationism*, by Tim M. Berra, to be published on 21 September by Stanford University Press. Berra provides a layman's explanation of the theory of evolution and argues against teaching creationism in US schools. Price is hbk \$29.50, pbk \$7.95. Barry Price, in *The Creation Science Controversy*, assesses the creationists' case and tells the story of the spread of these fundamentalist views in Australia. Published in Australia by Millenium, price A\$19.95 (pbk), the book will be issued in the United States and Canada in September. □

# First greening of Brazil

Alun Anderson

**The Burning Season. The Murder of Chico Mendes and the Fight for the Amazon Rain Forest.** By Andrew Revkin. Houghton Mifflin: 1990. Pp. 317. \$19.95. Published in the United Kingdom by Collins. Hbk £16, pbk £9.99.

CHICO Mendes was gunned down in the backyard of his home in the little Amazonian town of Xapuri just a couple of days before Christmas in 1988. He was on his way to wash in the bathhouse, towel over one shoulder, when he was hit by a shotgun blast. He staggered back into his house, where his two guards, his wife Ilzamar and his two small children were waiting, and died.

Andrew Revkin begins his tale with this murder, but the book is not out to solve a mystery. The murderers are known and, as the local Xapuri prosecutor later put it, "The fact is, everybody knew Chico Mendes was going to be killed sooner or later." Revkin's tale is a deeper one, of how Mendes came to be at this place at this time — the story of the personal transformations that took him from his childhood in a poor, uneducated rubber-tapping family, to his final years as an international spokesman for the preservation of the rainforest.

That transformation was never quite complete. Although his murder provoked an immediate international outcry and more than 1,000 people reached Xapuri in time for his funeral (no mean feat, as it is a twelve-hour air and road journey even from São Paulo), people in Xapuri still saw him not as the international figure he had become, but as the union leader he had been, fighting for the rights of poor rubber tappers. The attention paid to his death was not fully comprehensible: "Even when Jesus Christ died, there wasn't as much publicity," complained Sebastiao Alves da Silva, the patriarch of the Alves clan that is alleged to have carried out the murder. After all, Mendes was the forty-ninth rural activist to be killed in the Amazon that year.

Two things had made Mendes special to those in the developed world. First, he was Brazilian, and allies within Brazil for the environmental movement were in short supply. The second, and more enduring reason, was that the idea of the "extractive reserve" grew out of Mendes' experience as a rubber tapper. The traditional life of the rubber tapper — harvesting latex and Brazil nuts from several hundred hectares of land — did little harm to the forest. As a union leader, Mendes' first priority was to protect these peoples' livelihood from cattle ranchers. But as an environmentalist, Mendes saw their way of life in a more general equation — that it is rational to save the rainforest because it will be worth more if it can be harvested sustainably. That idea was quickly to seize the

imagination of the people in the developed countries, combining as it did both the romance of a life lived in harmony with nature, and the comfortable thought that conservation could be the best way to turn a profit — ecology could almost be a branch of economics. But much had to happen before Mendes found himself in Washington, describing the benefits of extractive reserves to congressmen.

Revkin tells us how the first of Mendes'

Mark Edwards

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

Chico Mendes' widow outside her house in Brazil.

transformations came as a boy. His family, like that of the other rubber trappers, was trapped by debt, by the constant need to repay the "advances" they received from middlemen who controlled the rubber trade. Into this unchanging world came Euclides Fernandes Távora, university-educated and a communist, who had escaped from an island penal colony.

Through Távora, Mendes learnt to read newspapers (there were no schools) and to understand the politics of the left. By the 1960s, he was working with the newly emerging unions, fighting debt slavery. A decade later it was the new enemy, the road cutting through the jungle from Rondonia that brought with it the cattle ranchers and hired gunmen sent in to "clean" the forest.

Revkin tells a fine tale of meetings with remarkable men and women, as Mendes emerges first as a national union leader and, soon after, as an international spokesman. The now standard chapters on the diversity of the rainforest and the global consequences of its destruction are

there too, but what puts the book in a category of excellence of its own is that it sees the problem of the Amazon as also the problem of the people who live and work in the Amazon.

By a miracle of publishing, Revkin's book finishes with events as recent as March 1990, fifteen months after Mendes' death. Mendes' murderers have still not been convicted but two important things have happened. First, Brazil has a new president in Fernando Collor de Mello, and a new secretary of the environment in José Lutzenberger, the country's most outspoken conservationist and an old friend of Mendes. Second, several large areas of forest have been designated as extractive reserves.

Both are successes that Mendes helped set in motion. But a big question remains over the long-term fate of Mendes' rubber tappers and the idea of extractive reserves. In passing, Revkin twice mentions that the price of rubber in Brazil is kept artificially high. He explains that native rubber collecting was almost wiped out at the beginning of the century when the British set up vast plantations in Malaysia, free from the pests that made them impossible to manage in Brazil. But Revkin does not go on to ask what this really means.

Unfortunately, the long-term outlook for extractive reserves is not good. Despite attempts to increase the number of products that can be harvested, extractive reserves seem unlikely ever to provide more than a subsistence income. Just as with rubber, if any plant is found to be especially profitable, then someone will find a way of cultivating it intensively. As Brazil develops — or if subsidies are removed — the way of life of the rubber tappers that Mendes fought so hard to protect may pass into history. The difference Mendes made, though, is that he contributed enormously to the first greening of Brazil's politics and he gave the rubber tappers a chance, through the new extractive reserves, to decide their own fate. □

Alun Anderson is Washington Editor of *Nature*.

■ *The World is Burning. Murder in the Rainforest — The Tragedy of Chico Mendes*, by Alex Shoumatoff, will be published in October by Chatto & Windus, price £15.

■ *Dam the Rivers, Damn the People*, a new Earthscan/WWF paperback by Barbara J. Cummings, is about development and resistance in Amazonian Brazil. Price is £5.95. Another new Earthscan paperback is *Our Common Seas: Coasts in Crisis*, in which Don Hinrichsen analyses the ecosystems being destroyed by industrial pollution. Price is £6.95.

■ *The Times Guide to the Environment* is published today. The author, Struan Simpson, aims to provide a comprehensive handbook to 'green' issues. Price for this paperback is £7.95. □



## In the beginning

Gerald F. Joyce

**Information and the Origin of Life.** By Bernd-Olaf Küppers. MIT Press: 1990. Pp. 215. £20.25, \$22.25.

MOLECULAR biology has been remarkably successful in following a reductionist programme aimed at providing a detailed understanding of biological phenomena in terms of the laws of physics and chemistry. It is natural to wonder to what degree this programme can be extended toward understanding the origin of life itself. All of biology rests on the foundation of neodarwinism, drawing on the principles of population biology and molecular genetics. A 'molecular-darwinistic' approach, developed by Manfred Eigen and others over the past 20 years, has been used successfully to extend neodarwinism to the level of a population of informational macromolecules. In his scholarly and intensely thought-provoking book, Küppers presents the case for molecular darwinism as he examines the philosophical underpinnings of the problem of the origin of biological information.

Küppers considers three possible explanations for the origin of biological information. The chance hypothesis, which he attributes to Jacques Monod, suggests that the specific sequence of the first genetic molecule was the result of a purely random process on the primitive Earth. Thus the origin of life was a unique and entirely unpredictable event. Küppers rejects this hypothesis on statistical grounds and, furthermore, applies algorithmic information theory to show that the chance hypothesis is inherently unprovable. The teleological approach, attributable to Michael Polanyi and others, argues that living systems possess a higher regulatory principle that goes beyond the laws of physics and chemistry and allows order to emerge against the universal tendency toward maximum entropy. Küppers rejects this hypothesis as well, pointing to the results of nonequilibrium thermodynamics which explain how order can emerge in an open system not at equilibrium. Again he turns to algorithmic information theory, this time to show that the teleological approach is inherently unfalsifiable.

Between the chance hypothesis and the teleological approach lies the molecular-darwinistic approach. According to this model, the origin of biological information results from the interplay between purely random mutational events and the higher-order principle of natural selection. Küppers firmly rejects any notion that natural selection is a special property of living systems, noting that while "natural selection in itself is not a physical law . . . neither is it an irreducible property of

living systems . . . rather, it results as a direct consequence of physical laws" (page 147). He describes in precise terms, following the work of Eigen, how selection can be characterized by an optimization principle that is a direct consequence of the behaviour of a population of self-replicating macromolecules that have different rates of production. Natural selection, therefore, does not succumb to the tautology 'survival of the survivor', but rather is a reflection of the quantifiable behaviour of self-replicating entities.

Küppers' refutation of the often-repeated statement that biological organisms can be understood only through a holistic approach that recognizes that "the whole is greater than the sum of the parts" is a treat for those who reject such fuzzy-headed thinking. Although he acknowledges that the methodology of the holistic approach often has heuristic value, he cautions against allowing "methodological mysticism to creep back into biology, at a time when this had only just begun to be eliminated by the development of molecular biology" (page 175).

A reductionist to the core, Küppers nonetheless has trouble establishing criteria to distinguish the living from the nonliving. Metabolism, self-reproduction and mutability are obviously necessary attributes, but are they sufficient to exclude inanimate structures such as crystals and viruses? Küppers adopts the uncomfortable position that "the transition from nonliving to living is a continuous one" (page 133). To him, a popula-

tion of self-replicating nucleic acids undergoing darwinian evolution lies in the realm of the prebiotic. What is lacking, in his view, is protein-based catalytic function. Others might argue that metabolism, self-reproduction and (heritable) mutability are sufficient criteria for life, thereby equating the origin of darwinian evolution, whatever its chemical basis, with the origin of life. But then the molecular-darwinistic approach would address the early evolution of life rather than the origin of life and one would be left without an explanation for the appearance of the first genetic molecule.

The only serious shortcoming of this book is that, despite a 1990 copyright, it is somewhat out of date. There are virtually no references to the scientific literature of the past 10 years, including no mention of catalytic RNA and no discussion of the recent work by Eigen on the movement of an evolving population through sequence space. The book was first published in German in 1986 and has apparently not been revised before publication of the English translation. Nonetheless, this is an important contribution to the philosophy of science. It merits serious consideration by biologists and by those interested in the question of how biology arose when once there was only physics and chemistry. □

Gerald F. Joyce is in the departments of Chemistry and of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037, USA.



This ocelot, *Felis pardalis*, is one of 36 species of wild cats in an "action plan for species conservation". From *Conserving the World's Biological Diversity* by J. A. McNeely et al., published by the World Bank. The authors explain the importance of diversity and call for action from businesses and institutions to preserve biological resources. Price is \$14.95 (pbk). (Photo by A. Young.) □

# The extragalactic Universe: an alternative view

H. C. Arp, G. Burbidge, F. Hoyle, J. V. Narlikar & N. C. Wickramasinghe

We discuss evidence to show that the generally accepted view of the Big Bang model for the origin of the Universe is unsatisfactory. We suggest an alternative model that satisfies the constraints better.

OVER the past twenty years, studies of the extragalactic Universe have been carried out by many astrophysicists — this is one of the most popular and spectacular branches of observational and theoretical astrophysics. In this period, a framework of belief has grown up in such a way that the observational phenomena have been fitted into a picture whose basic structure we now describe.

It is generally supposed that we live in an expanding Universe that originated in a hot Big Bang and that has an age between  $10$  and  $20 \times 10^9$  yr. The basis for belief in this cosmological model is the existence of the Hubble law for galaxies, which is well established for normal galaxies up to a redshift of about  $z \approx 0.5$ , the discovery of the microwave background radiation by Penzias and Wilson in 1965, together with the evidence that it was of black-body form, and the demonstration that the observed abundances of the light nuclei D,  $^3\text{He}$ ,  $^4\text{He}$  and  $^7\text{Li}$  could be explained in terms of early (Big Bang) nucleosynthesis. More recently, the popularity of the inflationary model and the conviction of many that the Universe has a density equal to the closure density have been amplified, not by observational evidence, but by the speculations of particle physicists eager to lead in the studies of the early Universe.

Parallel with these developments (and even earlier), studies of discrete extragalactic sources of non-thermal emission have contributed to the overall cosmological picture. First came the identification of the powerful radiogalaxies and the general phenomenon of radioemission. This was followed in the early 1960s by the discovery through radio techniques of the quasi-stellar objects (QSOs). It soon became clear that these phenomena were related to similar effects in the nuclei of galaxies — Seyfert nuclei — and in recent years there has been an irresistible urge to lump all these phenomena together under the heading of active galactic nuclei (AGN).

The characteristics of all of these discrete objects are the nonthermal character of their continuum spectra, the very strong and broad emission lines in the spectra, their variability in flux in many wavelengths ranges and, most important for their impact on extragalactic astronomy, the wide range in redshifts up to values of  $z > 4$ .

From the earliest days of the studies of radiogalaxies and, separately, of QSOs, it has been claimed that, provided their red-

shifts arise from the expansion of the Universe, these discrete sources show evidence for strong evolution in time, in their density and other physical characteristics. Thus the idea has developed that not only is the Universe as a whole evolving as it expands, but that populations of the discrete objects out of which the Universe is made are also evolving.

In what follows we shall argue that:

- The currently popular cosmological model is subject to many doubts based on observational data which suggest that, perhaps, there never was a Big Bang.

- The observational evidence concerning non-thermal objects with large redshifts leads inescapably to the conclusion that these redshifts are largely intrinsic in origin. This means that these objects do not lie at large cosmological distances, but in general that they lie much closer with  $z \leq 0.1$ . This in turn means that there is no evidence for evolution in the discrete objects and, as predicted by Ambartsumian<sup>1</sup> about 30 years ago, in these objects we are witnessing creation events involving the ejection of new matter from the nuclei of galaxies.

We discuss these two topics in reverse order.

## Nonthermal sources

For some years after Hubble demonstrated the law of the redshifts and tied it to the expanding Universe, attempts were made to reinterpret the redshifts as resulting from other causes — for example, the tired light phenomenon — but gradually these ideas died away.

Alternative explanations of the redshifts were not seriously attempted again until the discovery of large redshifts of the QSOs. Terrell<sup>2</sup> proposed that all these objects were ejected from the galactic centre and had passed us. By 1966, two arguments were going on. On the one hand, there was apparently no correlation between redshifts and apparent magnitudes for the QSOs. By various types of binning, the enormous scatter in the Hubble diagram could be reduced, but in no case did the correct slope of 5 naturally emerge<sup>3,4</sup>. Thus there was no real evidence for cosmological redshifts. On the other hand, we had discovered the severe problems imposed by the rapid variability of the QSOs; for we had to understand how the very high luminosity of a source can come from a very compact region whose size is set by the timescale of variability. F.H., G.B. and Sargent<sup>5</sup> discussed this

so-called Compton paradox and pointed out that it could be avoided if the objects were closer than was indicated by their redshifts, thus reducing the luminosities and hence their photon fluxes. F.H. and G.B.<sup>7</sup> then argued that a case could be made for bringing all the QSOs much closer than was indicated by their redshifts. Later, the discovery of apparent superluminal separation of compact radio components in a few QSOs demanded special geometrical assumptions with highly relativistic bulk motions to explain the observations as illusions<sup>8</sup>. Such contrived assumptions are not needed if the QSOs are much closer than their redshift distances.

Yet there was still no direct evidence for the distances. This began to be collected only when H.C.A. started to look at the associations between peculiar galaxies and radio sources<sup>9</sup>, and when G.B. *et al.*<sup>10</sup> studied a well-defined sample — the 3CR QSOs and their proximity to bright galaxies in the Shapley-Ames catalogue. Over the next 15 years, more data have accumulated, and several samples have been analysed in detail. It soon became clear from statistical arguments alone that there are more QSOs close to bright galaxies than there should be by accident, thus implying that QSOs with high redshifts are physically associated with galaxies of low redshifts. But it was often argued, by those who were not prepared to believe in physical associations, that proper statistical methods were not being used.

In 1971, H.C.A.<sup>11</sup> found the remarkable optical bridge between Mk 205 ( $z = 0.07$ ) and NGC 4319 ( $z = 0.0057$ ), about  $40''$  away. Controversy continued about the reality of this feature until the definitive study by Sulentic and H.C.A.<sup>12,13</sup>. In 1979, G.B.<sup>14</sup> analysed all QSOs found near bright galaxies and showed that, within separations of  $3''$ , the results were highly significant. Other landmarks along the way included the discovery of three QSOs within  $2''$  of the centre of NGC 1073 (ref. 15), three QSOs within  $2''$  of NGC 3842 (ref. 16) and other multiple systems investigated by H.C.A. The strong evidence up to 1987 is contained in H.C.A.'s book<sup>17</sup>.

Most recently, G.B. *et al.*<sup>18</sup> have compiled nearly 500 QSO-galaxy pairs mostly lying within  $10''$  of each other. This includes both QSOs near bright galaxies found by a computer search and QSOs near faint galaxies reported in the literature. Detailed statistical analyses of these large samples leads to the conclusion



that the QSOs with large redshifts are physically associated with galaxies with lesser redshifts, both for faint and bright galaxies.

In the past few years, work by other groups (see refs 19–21) has produced further samples of QSO galaxy associations far above the level expected if these were chance configurations. In fact, it is often the case that when a QSO is identified, one or more galaxies are found to lie nearby. So far, nearly 4,500 QSOs with redshifts have been catalogued<sup>22</sup>. From the counts of galaxies, the sky-surface density of galaxies brighter than  $15^m$  is about 0.25 per square degree. Thus, on a random distribution, the probability of finding such a galaxy within  $2'$  of a QSO is  $\approx 4 \times 10^{-4}$ . Thus, out of all the catalogued QSOs, we would expect about two pairs with separation  $2'$ . This is at least an order of magnitude less than the number found. We therefore consider that the evidence for the physical association of QSOs of high redshift with galaxies having much smaller redshifts is by now overwhelming.

The only way of understanding the observational evidence that would allow galaxies and QSOs to lie at different distances is the proposal<sup>23</sup> that the QSO is gravitationally microlensed by a star in the halo of the foreground galaxy. For some years, this was considered to be a possible solution, and it is enlightening to note that Stocke *et al.*<sup>19</sup> were prepared to believe in their own result only because they thought that it could be explained by gravitational microlensing. We quote from their paper:

For several years now there has been some evidence for a slight statistical excess of QSOs projected on the sky near bright galaxies. Over most of the history of these observations, the sole interpretation has been that these QSOs were physically associated with the bright galaxies near them and that the QSO redshifts were, therefore, "non-cosmological". Because this interpre-

tation was contrary to the well-accepted cosmological redshift hypothesis, this evidence was largely dismissed on the grounds that the data were either statistical fluctuations or were due to an improper statistical analysis. More recently, another interpretation has arisen following the discovery of multi-image QSOs in which QSOs near foreground galaxies are brightened by the gravitational effects of individual stars in these galaxies near to our line of sight, so-called minilensing or microlensing. This new interpretation is consistent with the cosmological redshift hypothesis and has revitalized interest in the observational data on QSO-galaxy associations.

In fact, this explanation will not work<sup>24–26</sup>. Essentially, the reasons are that the observed luminosity function for QSOs shows that there are not enough faint ones to be amplified by microlensing, and in halos there are not enough mass points. The following argument by F.H. demonstrates this:

(1) A galaxy with its mass distributed over a radius  $\sim 10$  kpc does not produce QSO amplification unless the mass exceeds about  $5 \times 10^{11}$  solar masses ( $M_\odot$ ). Divided into many smaller masses of stellar order, however, amplification by microlensing can occur, even though the total mass is less than  $5 \times 10^{11} M_\odot$ .

(2) When a QSO assumed to be cosmologically distant is known to be projected against a nearby galaxy, the possibility of the received radiation being magnified by a factor  $Q$  above what it would otherwise have been is  $\sim 2 \times 10^{-2} Q^{-2}$ . For example, the chance of a twenty-second magnitude QSO being magnified so as to appear to the observer as a seventeenth magnitude QSO is  $\sim 2 \times 10^{-6}$ . There would need to be  $5 \times 10^5$  times more QSOs at the twenty-second magnitude than at the seventeenth for the number observed at the seventeenth to be significantly increased by microlensing. This is a vast excess, several hundred times at least, over the number of QSOs observed at the twenty-second

magnitude.

(3) Quite apart from the small coefficient of  $\sim 2 \times 10^{-2}$ , the number of faint QSOs would need to increase more rapidly than  $2.5 \log_{10} Q$  for magnification by microlensing to be significant. Although such an increase with magnitude holds between the fifteenth and seventeenth magnitudes, the increase at still fainter magnitudes is much less than  $2.5 \log_{10} Q$ .

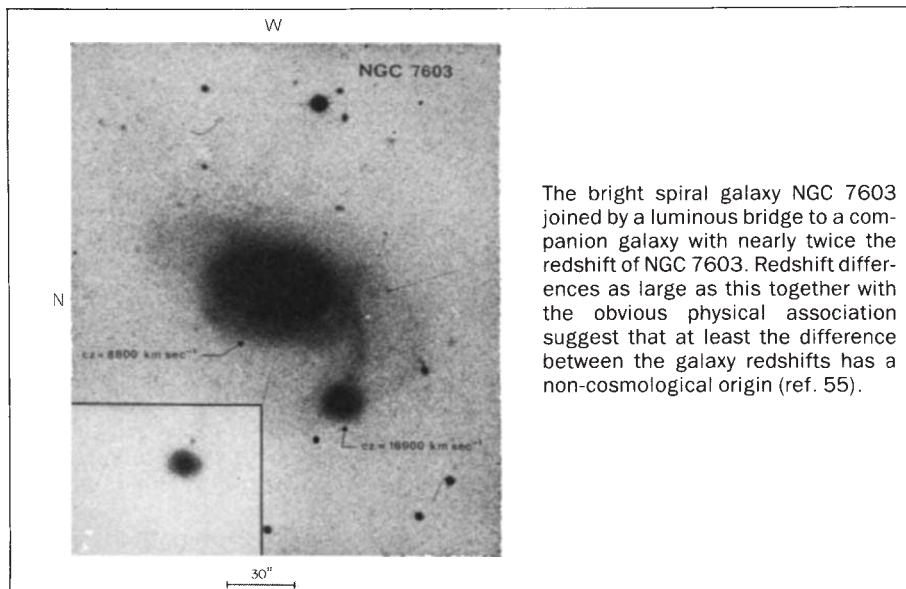
(4) Changing nearby galaxies to distant galaxies does not increase the small normalizing coefficient  $2 \times 10^{-2}$  at all appreciably. Microlensing by faint galaxies is just as unimportant as microlensing by bright galaxies.

(5) Even if the microlensing had worked, it would not apply to the observed QSO-galaxy pairs where the separation is of the order of two to three times a typical galaxy radius.

It is widely believed that the blobby structure of the QSO image near the centre of 2237+305 is evidence of a gravitational lens<sup>27–29</sup>, but the chance of this is  $\sim 2 \times 10^{-6}$ , which is appreciably less than the chance,  $\sim 5 \times 10^{-5}$ , of the QSO being accidentally projected against the associated galaxy; the lens suggestion is highly unlikely. The blobby structure of the QSO image can rather be interpreted as evidence of emission from the central regions of a strong gravitational well. Light emitted at angles to the radial direction is curved by  $\pi/2$  or more, so that to an external observer the QSO light appears as a halo encompassing the entire well. The halo has a blobby structure simply because the sources within the well happen to be irregularly distributed.

Thus it seems that in some way the QSOs are physically connected to the normal galaxies both for  $z_Q \gg z_G$  and where  $z_G \approx z_Q$ . Physically, this means that QSOs often appear and are formed in comparatively dense regions of galaxies. This result can be extended to explain the high frequency of the presence of absorption in the spectra of QSOs, usually interpreted as meaning that there is much absorbing matter ejected from the QSO and/or there are many absorbing clouds along the line of sight. Since the discovery of this effect, it has been realized that the amount of absorption requires a cross-section of absorbing matter greater than expected from normal galaxy halos by a factor of five or more<sup>30</sup>. This has been attributed to the presence of more absorbing matter of high redshift than is present locally<sup>31</sup>. But it is easily seen that it is better interpreted in terms of our newly found rule: it is due to abnormal amounts of matter with very different redshifts all lying near the QSO. It has also been reported that there are more normal galaxies than expected by chance near to powerful 3CR radiogalaxies<sup>32</sup>. This suggests that the rule can also be extended to such objects.

Our conclusion, therefore, is that the



evidence for the effect is overwhelming and a general rule can be stated. More galaxies than expected by chance are found very close to QSOs and other active nuclei. This conclusion applies to associations both between QSOs and bright galaxies and between QSOs and faint galaxies; it also applies both in situations in which  $z_Q \approx z_G$  (refs 33,34) and when  $z_Q \gg z_G$ .

### Cosmic microwave background

The strongest evidence cited in favour of the hot origin of the Universe is the observed radiation background<sup>35</sup>. The recent announcement by COBE<sup>36</sup> that the spectrum of the radiation is planckian to a high degree of approximation will be interpreted by many as a confirmation of their belief in the correctness of the Big Bang model. But there is bad news for the Big Bang as well as good. The effective temperature of the radiation is remarkably constant with respect to direction, both on small and large angular scales. Thus, on a scale of a few arcminutes there is constancy in the temperature to within about 1 part in 50,000 (ref. 37).

But, in the Big Bang model, an early planckian spectrum will be distorted by subsequent events of a non-thermodynamic character, of which the condensation of galaxies, of clusters of galaxies and, perhaps, of a still larger cellular structure on a present-day distance of about 100 Mpc, are the most frequently discussed examples. The microwave background has no imprints to mark the occurrence of such events, contradicting the theoretical expectations of a decade ago and causing theoreticians in recent years to search for variants of the Big Bang that avoid a confrontation with observation on this point. Our opinion is that avoiding confrontation with observation is not the hallmark of a good theory.

The Big Bang model offers a Universe created in a smooth featureless condition, out of which a highly structured Universe is nevertheless supposed to have evolved. Numerous attempts have been made to explain how this miracle is supposed to have happened. They have two features in common, one a retreat into the highest flights of physics and the other an unsatisfactory absence of the immense detail that would be required to support them in a proper manner, from which we suspect the attempts to be little more than ingenious handwaving. Perhaps this is why they are called 'scenarios'.

The root of the matter, it seems to us, is one of time-ordering. In the Big Bang model, the microwave background came first and the galaxies second, whereas the observations suggest (almost to the point of compelling) the opposite. For the microwave background to come second, it is necessary that some form of particle should be widespread extragalactically,

either at present or in the relatively recent past, with the property of being strongly absorptive of microwaves, yet of being almost translucent in both the visible and longer radio wave regions of the spectrum. We can ask three questions:

- Do such particles exist as a theoretical possibility?
- Is there laboratory evidence for their existence?
- Is there astronomical evidence of their presence?

The answer to the first question is certainly yes. The optical constants of metals are such that the absorptivity of thin metallic needles, determined by a Mie-type calculation<sup>38</sup>, is 100 times greater at a wavelength of 1 mm than it is at 5,000 Å. This is at room temperature. At cryogenic temperatures, because of the lowering of metallic resistivity with temperature, the factor is increased from 100 to 1,000. Thus extragalactic space could be exceedingly opaque to microwaves and still exceedingly translucent in the visible. Depending on their diameters and lengths, metallic whiskers ultimately lose their high absorptivity at sufficiently long wavelengths. A whisker of length 1 mm would, for example, lose its high absorptivity at a radio wavelength of about 10 cm.

The answer to the second question is also yes. Laboratory experiments (ref. 39, D.V. Gal'tsov, personal communication) show that slowly cooled metallic vapours condense into whiskers typically with radii of  $\sim 10^{-6}$  cm and lengths of about 1 mm. The experiments are explained in a scheme where the first nucleation occurs into liquid droplets containing a few thousand atoms. When sudden crystallization eventually occurs to the solid state, the considerable energy released causes the resulting crystal to possess dislocations. One particular dislocation, the screw dislocation, has the property of causing subsequent growth to occur linearly, whereas other dislocations lead to more or less spherically growing particles. Now it can easily be shown that, whereas the radius of a spherically growing particle increases proportionally to the time, the length of a linearly growing particle increases exponentially. Thus those nucleations that happen to develop the linearly growing property soon come to dominate. Because of their exponential growth they largely consume the available supply of vapour.

There are two lines of astronomical evidence pointing to an affirmative answer also to the third question, both connected with supernovae. First, because of their synthesis of metals, one would expect metallic whiskers — iron whiskers especially — to condense in expanding ejecta from supernovae, just as they do in the laboratory. Radiation from the pulsar in the Crab Nebula has been found<sup>40</sup> to be drastically weakened over the wavelength

range from about 30  $\mu$ m to about 10 cm, just the range in which the absorptivity of iron whiskers would be strong. Apart from the unlikely possibility that the absence of this band of radiation is an intrinsic property of the pulsar itself, the natural explanation is that whiskers in the supernova ejecta, still surrounding the pulsar (but not surrounding the outer regions of the Crab) are absorbing the missing radiation.

Second, there is a remarkable correlation between the radio luminosities of galaxies and their luminosities in the far infrared ( $\sim 100 \mu$ m), a correlation maintained over some four orders of magnitude in the intrinsic emissions<sup>41</sup>. The radio emissions come via the synchrotron process from high-energy electrons derived from pulsars. The supernovae that provide the pulsars would also provide quantities of iron whiskers. The greater the frequency of supernovae the larger the number of pulsars and the greater the quantity of whiskers, thereby establishing the required correlation in what otherwise would seem to be an inexplicable situation.

The absorptivity of iron whiskers is so large in the far infrared that whiskers in a local infrared radiation field experience strong outward radiation pressure. Using known energy densities for such fields it can be calculated that whiskers could attain speeds of emission into extragalactic space as high as 10,000 km s<sup>-1</sup>, sufficient to carry them from their parent galaxies to a distance of  $\sim 100$  Mpc in a time  $H^{-1}$ . Viewed cosmologically, therefore, the distribution of whiskers in extragalactic space could be much more uniform than the galaxies producing them. In effect, the whisker production of  $10^6$  galaxies becomes averaged together in a Hubble time.

In relation to this, note that once a radiation field has been thermalized, the distribution of the thermalizing agent is irrelevant. The thermalizing agents are not required to be smooth. It is the energy density of the radiation that is required to be smooth. In this regard, suppose a microwave quantum to have a mean free path  $\lambda$  before being absorbed and re-emitted. In a time  $H^{-1}$ , radiation then has a diffusion distance  $(\lambda c H^{-1})^{1/2}$  which is  $\approx 0.1 c H^{-1}$  for  $\lambda = 10^{-2} c H^{-1}$ , implying that in a time  $H^{-1}$ , fluctuations in the radiation field would be greatly smoothed, even for a cosmological optical depth for microwaves as large as 100.

Consequently, there is a double tendency towards smoothness, a tendency in the thermalizing agent and a strong tendency in the radiation field. If the cosmological model were also steady-state, one could expect this double tendency, combined with an ongoing cosmological situation, to have produced the high degree of smoothness that observations show the micro-



wave background to possess.

Fast-moving whiskers are exposed to sputtering effects, which are at a maximum for speeds of a few hundred  $\text{km s}^{-1}$  — the effects become less important at higher speeds because impacting micro-particles have less time to impart momentum to the material of a whisker. A critic, anxious to find a flaw in the above argument, could therefore argue that sputtering would destroy whiskers as fast as they were produced. As well as overlooking the astronomical evidence cited above, such a criticism also misses the important point that, because of their high emissivity in the far infrared, whiskers tend to stay very cool. For the most part their temperatures cannot much exceed that of the microwave background itself, even for whiskers in the vicinity of a galaxy. Foreign molecules therefore condense readily on their surfaces, and it would largely be such foreign molecules that would suffer evaporation by sputtering. All solid substances except  $\text{H}_2$  condense at a temperature of 2.7 K.

The commonsense inference from the planckian nature of the spectrum of the microwave background and from the smoothness of the background is that, so far as microwaves are concerned, we are living in a fog and that the fog is relatively local. A man who falls asleep on the top of a mountain and who wakes in a fog does not think he is looking at the origin of the Universe. He thinks he is in a fog.

### Expectations of the steady state

**Young Galaxies.** In the days when both Big Bang and steady-state cosmologies were being widely debated it was often pointed out that if the steady-state theory was correct, we would expect to see at all redshifts a small fraction of galaxies that were genuinely young in evolutionary terms. For example, in 1963 it was pointed out that NGC 2444-45 might be such a system<sup>42</sup> as it appeared to contain nothing but gas and hot O and B stars. It was immediately argued that the continuum of this remarkable hot spot galaxy had colours suggesting an intermediate age stellar population<sup>43</sup>.

Because O and B stars have ages of at most a few tens of millions of years, such attempts to identify young galaxies were, from the point of view of the steady-state theory, unnecessarily rigorous, as galaxies with ages  $\sim 3 \times 10^9$  yr would still be 'young' according to the theory. It is widely assumed by cosmologists that essentially all galaxies have closely the same age. But there is no proof of this, as the integrated colours of distant stellar populations are insensitive to age once the age increases above  $\sim 10^9$  years. It is perhaps conceivable that with the great improvements of observational technique now in the offing, subtle differences of integrated colours may yield information on the ages of

galaxies. Significant variations might well prove to be one of the shocks that cosmologists will experience in the future.

Despite the over-rigorousness of the young-star criterion from the point of view of the steady-state theory, it is worth noting that there are now many known examples of 'star burst' galaxies dominated by recently formed stars. Recent infrared surveys with the IRAS satellite reveal considerable numbers of galaxies which are indicated to be young in evolutionary terms<sup>44</sup> as they appear to contain little else but atomic and molecular gas and high-mass young stars. The number of such galaxies that may be very young — objects such as Arp 220, NGC 6240 are some of the most luminous<sup>45</sup> — is small but they are nearby and fit well with predictions of the steady-state model. In the framework of the Big Bang cosmology they are not understood at all and would have to be protogalaxies long delayed in their collapse. But protogalaxy clouds of gas in the process of collapsing are not observed even after extensive search and therefore the presence of young galaxies furnishes yet another objection to the Big Bang theory.

**The age of the Universe.** At the time the steady-state theory was proposed (1948) the age of the Universe appeared to be less than the ages of objects in it and this was very much an argument in its favour. Depending on the value chosen for the Hubble constant and other factors, the age discrepancy problem may be back with us. For  $H = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , the lowest value currently favoured, the typical age would be  $2/3 H = 13 \times 10^9$  yr. For a value as large as  $80 \text{ km s}^{-1}$ ,  $2/3 H = 8 \times 10^9$  yr. Values for the ages of the oldest stars are currently about  $13\text{--}17 \times 10^9$  yr (ref. 46), and the age of the elements  $\sim 12\text{--}16 \times 10^9$  yr (ref. 47). Thus, for most of the range of allowed values the age of the oldest stars is greater than the age of the Universe. For some reason it is not being discussed, but in terms of numerical factors the problem for the Big Bang is back again.

The resolution of the age problem within the conventional framework lies only in abandoning the inflationary model or in robbing the model of its only important good point by finely tuning a cosmological constant. This constant has been in and out of cosmology ever since Einstein introduced it in 1917. It is brought in whenever it is felt that observations warrant it, only to be abandoned when theoreticians find that they can do without it.

**Evolution of discrete objects.** One of the arguments that has continually been used to attack the steady-state cosmology is that counts of discrete sources show strong evolutionary effects. It is obvious from what we have discussed above that if these discrete sources are not at great distances

this argument disappears.

It is also important to point out that a detailed study of the  $\log N\text{--}\log S$  relation was recently carried out for the 3CR radiogalaxies, a sample in which practically all the redshifts are now measured. It was shown that, even on the assumption that these redshifts are cosmological in origin, this most important sample of radio sources shows a  $\log N\text{--}\log S$  distribution entirely compatible with the steady-state model<sup>48,49</sup>.

Moreover, the discovery of the highest redshift quasar at  $z = 4.7$  was recently announced (ref. 50). Even though it must be far back near the age of formation of galaxies on the Big Bang hypothesis, it shows no spectroscopic differences from lower redshift, presumably much nearer quasars. This is direct evidence against evolution of properties of the universe over a major portion of its lifetime.

We conclude this section by emphasizing that each of these arguments which was used in the debate some twenty years ago concerning the viability of the steady-state cosmology is still with us and on balance the steady-state model is favoured.

### Alternative theories

The above discussion clearly indicates that the present evidence does not warrant an implicit belief in the standard hot Big Bang picture. What are the alternative ideas? We do not claim to present here 'the' alternative to the standard picture. Rather we draw the attention of the reader (and the prospective research worker in the field) to some ideas that have already addressed these questions at least partly.

It is important to decide whether the issue of non-cosmological redshifts discussed in the earlier sections are of chief importance to cosmology itself. But that there is a serious problem (as opposed to a cosmological problem) we are not in doubt. There is simply too much evidence to be dismissed as coincidental. Even in the short time since we wrote this article, evidence has appeared<sup>51</sup> to show that the prototype quasar 3C 273 is situated in the Virgo cloud of galaxies at a cosmological redshift of only  $1,170 \text{ km s}^{-1}$ , far less than the redshift of 3C 273 0.16 ( $0.16c \approx 48,000 \text{ km s}^{-1}$ ).

Adherents of the popular cosmological interpretation of QSO redshifts dismiss such situations as coincidental, even though the probability of coincidence is very low. For the situation concerning 3C 273, the probability is  $\sim 10^{-3}$ . Small but not as small as in some other cases. The psychology that permits such situations to be dismissed is that of mental compartmentalization. Stop thinking about the million-to-one chance of three QSOs lying within  $2'$  of NGC 3842, or about the 100,000 to 1 chance of a system like 2237 + 0305 being found, or the many other

examples we could mention, when thinking about 3C 273. And of course omit 3C 273 when thinking about any one of the others. And so on in turn down the list.

Turning now to cosmology, the Big Bang is a necessary consequence of Einstein's theory of relativity. Unlike all other parts of physics, the Einstein theory is not scale-invariant. But a gravitational theory can indeed be constructed possessing scale invariance<sup>32</sup>, and if one thinks that the whole of physics should be scale-invariant this formulation should be followed rather than Einstein's. The latter then appears as a special case of the more general theory. The general theory is wider than Einstein's in that it admits of two other cases. The other cases can be written mathematically in a form like Einstein's but with extra terms appearing in the field equations. One case is equivalent to introducing the so-called cosmological constant into the equations, which of course was done empirically by Einstein himself and used by later workers, notably by Lemaitre. It is this case that is used in modern inflationary cosmology.

The wider second possibility admits the creation of matter, which is to say it permits the trajectories of particles to be segments rather than paths extending to infinity in both the past and the future. Is this not just what the Big Bang model does, one can ask? Are the paths of particles not half-lines with a beginning at the origin of the Universe, and then extending infinitely into the future? The answer to these questions is that it is precisely because the Big Bang model seeks to bring segmented paths into a theory that does not admit them that the theory breaks down. In other words, the 'origin of the Universe' is not what is usually supposed, a true origin of the Universe. It is rather a breakdown of the theory itself.

The theory permitting segmented paths has no such breakdown. The full range of

mathematical solutions permitted by the wider theory has not yet been explored. In a first approximation there are two forms of solution, one the old steady-state solution, which had a high closure density analogous to that of the Einstein-de Sitter ( $q_0 = +0.5$ ) case among the Friedman models, and the other at a significantly lower density comparable to the mean density of baryonic matter that is normally observed. This second possibility is much harder to investigate because in a second approximation it is necessarily inhomogeneous, according very well with the observed distribution of galaxies on inhomogeneous scales up to at least 100 mpc. The difficulty for theoretical analysis is that the cosmological equations are necessarily of a partial character. Although the gravitational equations in their most general form are of course of a partial character, it is a feature of orthodox cosmology that approximations are used to simplify them into ordinary differential equations. It is the assumption of orthodox cosmology that nothing of major importance is lost in this simplification. It is a critical mathematical feature of the second possibility outlined above that it challenges this conventional assumption.

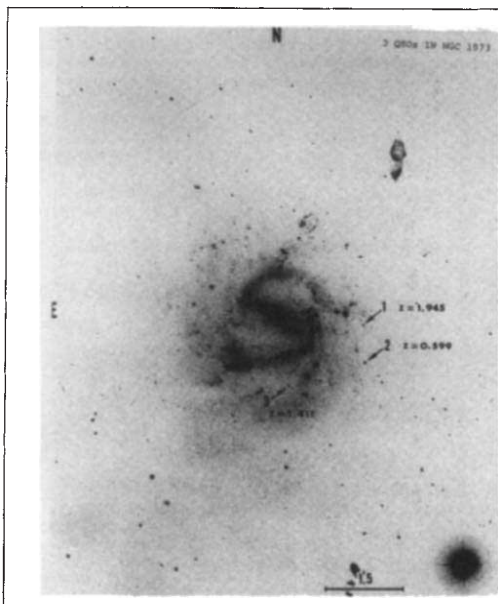
In a scale invariant gravitational theory it is possible to choose a particular conformal frame in which particle masses are everywhere the same, as they are normally supposed to be. Inhomogeneities then appear as gravitational potential wells, producing locally generated redshifts. Alternatively, a conformed frame can be chosen to remove the potential wells, at least to the extent of removing the Riemann scalar curvature. When this is done, the redshifts appear again as variations of particle masses. Thinking of such redshifts as 'anomalous', the anomalous values required by the second possibility of the previous paragraph appear to be about one-third of those arising from the

usual Hubble flow. This is on the average, and on the scale of the larger inhomogeneities, say 100 mpc. On smaller scales of the order of 1 mpc, and still more on the scale of QSOs, the anomalous redshifts required by theory could be appreciably larger than one-third of the Hubble flow. We think that it is in such terms that the facts concerning anomalous redshifts, discussed earlier in this article, are to be explained. This has been the mathematical basis underlying our position concerning those data. There are two ways to extend the existing position, one to continue to seek further connections with observation, the other to introduce real particles into the theory in place of classical particles.

Considering relations with observation further, and noting that the cosmological equations always have local solutions essentially identical to the main cosmological model, it is tempting to incorporate the Big Bang concept in the form of local events. Such local events could vary in scale from masses that have been associated with events in galactic nuclei, say  $10^8 M_\odot$ , up to the masses of whole galaxies, and thence to the masses of clusters and superclusters, the former being relatively common and the latter of a cosmological degree of rarity — that is, happening in any particular locality only once in a timescale of about  $H^{-1}$ .

It is important that, as the theory admits the creation of matter within its mathematical structure, creation events of this kind would be causal. They would be related to the Universe as a whole, mathematically through boundary conditions of the partial differential equations governing their internal properties and moments of occurrence. Otherwise the physics involved would be no different in principle, and for larger masses of the order of superclusters no different in detail from the situation in the supposed universal Big Bang. Associations between particle physics and the Big Bang translate, so far as physics is concerned, into similar associations at the large-scale end of creation when treated as a non Big Bang phenomenon.

It is commonly supposed that the so-called primordial abundances of D,  $^3\text{He}$ ,  $^4\text{He}$  and  $^7\text{Li}$  provide strong evidence for Big Bang cosmology<sup>33</sup>. But a particular value for the baryon-to-photon ratio needs to be assumed *ad hoc* to obtain the required abundances. A theory in which results are obtained only through *ad hoc* assumptions can hardly be considered to acquire much merit thereby. According to our point of view, on the other hand, the required abundances follow inescapably from the density-temperature relation that holds in all bodies with masses large compared with that of the Sun. Note that in the steady-state theory the Universe expands because of the creation of matter



Three QSOs lying within  $2''$  of the centre of the nearby spiral galaxy NGC 1073. The redshift of NGC 1073 is 0.004, or  $cz = 1,200 \text{ km s}^{-1}$ . However the QSOs have redshifts of 1.95, 0.60 and 1.41. Since the density of such objects over the whole sky is very small, no more than about 20 per square degree, the chance of finding three objects this close to the galaxy is exceedingly small, perhaps one in a million (ref. 16).



within it. The space-time structure of each new creation unit makes room for itself by shouldering aside the products of previously existing units, a process that requires each creation unit to expand with just sufficient energy to fit itself into the general Hubble flow of the Universe. This condition sets up the relation mentioned above between density and temperature, in which the density of matter is proportional to the cube of the temperature of the internal radiation field, the same relation as in Big Bang cosmology. But instead of the constant of proportionality being chosen arbitrarily to fit the required primordial abundances of D,  $^3\text{He}$ ,  $^4\text{He}$  and  $^7\text{Li}$ , the constant in the steady-state case is determined by the scale of the creation unit.

The scale of the creation unit so determined matches very well the size of superclusters of galaxies. The picture is of a Universe out to  $z = 1$  composed of about 1,000 major creation events, with creation occurring at cosmologically different epochs. Some among the 1,000 events will be relatively young, say  $0.3H^{-1}$ , others will be older, say  $4/3 H^{-1}$ , which is the age that seems necessary for 'our event' to accommodate comfortably the ages of the oldest stars in local galaxies ( $17 \times 10^9$  yr) to what seems to be the best determined value of  $H^{-1}$ . The need for us to live in the now-scattered residues from such an old event, rather than in the probably more numerous younger events, could well be anthropic. The need for human evolution to have taken place sets a lower limit to the age of our 'local' event. This lower limit may well be as large as  $4/3 H^{-1}$ .

There are several points of superiority of our view compared with the Big Bang hypothesis. One is that the baryon-to-photon ratio is set by the above considerations. Once the scale at the upper level of creation events has been set at  $0.1 H^{-1}$ , the baryon-to-photon ratio comes out necessarily as about  $3 \times 10^{10}$ , the best value for the synthesis of the light elements. In the Big Bang theory, on the other hand, this ratio has to be assumed *ad hoc*.

By having a wide spectrum in the scale of creation events (with those of smaller scale much more frequent than the larger scale of  $\sim 0.1H^{-1}$ ) there is a natural continuity between cosmology on the one hand, and the events associated with non-thermal emission — radio and X-ray sources — discrepant redshifts, and violent events on the scale of whole galaxies, which the observations clearly show to be intimately connected with one another.

The advantages of such a point of view can be measured both in details and in general principles. There has up to now been no really good stellar suggestion for the nature of the  $r$ -process in nucleosynthesis. Creation events for masses of  $\sim 10^8 M_{\odot}$ , however, provide a well-nigh perfect

process, both for the production of quantities of metals, but also for their driving by rapid neutron addition to the upper regions of the periodic table.

The conventional critic may argue that from the standpoint of economy of postulates the idea of 'many' creation events is a lot worse than the notion of the single creation event (the Big Bang). We disagree. The 'many' events in our alternative theory are potentially observable and satisfy the repeatability criterion of physical theories. The Big Bang satisfies neither of these requirements, and hence as a scientific hypothesis fails to compete with the alternative proposed here.

Cosmology is unique in science in that it is a very large intellectual edifice based on very few facts. The strong tendency is to replace a need for more facts by conformity, which is accorded the dubious role of supplying the element of certainty in people's minds that properly should only belong to science with far more extensive observational support. When new facts do come along, as we believe to be the case with anomalous redshifts, it is a serious misprision to ignore what is new on the grounds that the data do not fit established conformity. Certainty in science cannot be forthcoming from minimal positions such as those which currently exist in cosmology. It is true that when we come to astrophysics many more facts become available, and if the efforts of the past

quarter-century had succeeded in matching astrophysics to cosmology, the case would be much stronger. But this is precisely what has not happened. Astrophysics remains apart from cosmology, just as much as it was in 1965 when the microwave background was discovered. On this ground alone we think suspicions must be entertained that something in conformist cosmology is very wrong indeed.

As a general scientific principle, it is undesirable to depend crucially on what is unobservable to explain what is observable, as happens frequently in Big Bang cosmology. Geology progressed favourably from the time Hutton's principle of uniformity was adopted<sup>41</sup>, according to which everything in geology is to be explained by observable ongoing processes. One can suspect that cosmology and cosmogony would profit similarly from the adoption of the same principle, as the view proposed here does. □

H.C. Arp is at the Max-Planck Institut für Astrophysik, 8046 Garching bei München, FRG. G. Burbidge is at the Center for Astrophysics and Space Sciences, University of California, San Diego, La Jolla, California 92093, USA and acknowledges research support from NASA. F. Hoyle and N.C. Wickramasinghe are at the School of Mathematics, University of Wales College of Cardiff, Cardiff CF2 4AG, UK. J.V. Narlikar is at the Inter-University Centre for Astronomy and Astrophysics, Ganeshkhind, Pune 411007, India.

1. Ambartsumian, V.A. in *La Structure de l'Univers* (ed. Stoops, R.) (Solvay Conference Reports, Bruxelles, 1958).
2. Terrell, J. *Science* **145**, 918–919 (1964).
3. McCrea, W.H. *IAU Symp.* **44**, 282–286 (1972).
4. Bahcall, J.N. & Hills, R.E. *Astrophys. J.* **179**, 699–703 (1973).
5. Burbidge, G.R. & O'Dell, S.L. *Astrophys. J.* **183**, 759–766 (1973).
6. Hoyle, F., Burbidge, G. & Sargent, W.L.W. *Nature* **209**, 751–754 (1966).
7. Hoyle, F. & Burbidge, G. *Astrophys. J.* **144**, 534–552 (1966).
8. Zensus, J.A. & Pearson, J.T. (eds) *Superluminal Radio Sources* (Cambridge University Press, 1987).
9. Arp, H.C. *Astrophys. J.* **148**, 321–365 (1967).
10. Burbidge, E.M., Burbidge, G.R., Solomon, P.M. & Strittmatter, P.A. *Astrophys. J.* **170**, 233–240 (1971).
11. Arp, H. *Astrophys. Lett.* **9**, 1–4 (1971).
12. Sulentic, J.W. *Astrophys. J.* **265**, L49–L53 (1983).
13. Sulentic, J.W. & Arp, H.C. *Astrophys. J.* **319**, 693–708 (1987).
14. Burbidge, G. *Nature* **282**, 451–455 (1979).
15. Arp, H.C. & Sulentic, J.W. *Astrophys. J.* **229**, 496–502 (1979).
16. Arp, H.C. & Gavazzi, G. *Astr. Astrophys.* **139**, 240–242 (1984).
17. Arp, H. *Quasars, Redshifts and Controversies* (Interstellar Media, Berkeley, 1987).
18. Burbidge, G., Hewitt, A., Narlikar, J.V. & Das Gupta, P. *Astrophys. J. Suppl.* **74**, (in the press, November 1990).
19. Stocke, J.T., et al. *Astrophys. J.* **315**, L11–L15 (1987).
20. Webster, R.L., Hewitt, P.C., Harding, M.E. & Wegner, G.A. *Nature* **336**, 358–359 (1988).
21. Fugmann, W. *Astr. Astrophys.* **204**, 73–80 (1988).
22. Hewitt, A. & Burbidge, G. *Astrophys. J. Suppl.* **63**, 1–246 (1987); **69**, 1–63 (1989).
23. Canizares, C.R. *Nature* **291**, 620–624 (1981).
24. Arp, H.C. *Astr. Astrophys.* **229**, 93–98 (1990).
25. Linder, E.V. & Schneider, P. *Astr. Astrophys.* **204**, L8–L10 (1988).
26. Hogan, C.J., Narayan, R. & White, S.D.M. *Nature* **339**, 106 (1989).
27. Huchra, J. et al. *Astron. J.* **90**, 691–696 (1985).
28. Yee, H.K.C. *Astron. J.* **95**, 1331–1339 (1988).
29. De Robertis, M.M. & Yee, H.K.C. *Astrophys. J.* **332**, L49–L53 (1988).
30. Roberts, D.H. et al. *Astrophys. J.* **224**, 344–367 (1978).
31. Sargent, W.L.W. *IAU Symp.* **124**, 777–793 (1987).
32. Hammer, F. reported at *Workshop on Radio Galaxies* (University of California, Berkeley, 1989).
33. Gehren, T., Fried, J., Wehinger, P.A. & Wyckoff, S. *Astrophys. J.* **278**, 11–27 (1984).
34. Yee, H.K.C. *Astron. J.* **94**, 1461–1467 (1987).
35. Penzias, A.A. & Wilson, R.W. *Astrophys. J.* **142**, 419–421 (1965).
36. Mather, J.C. et al. *Astrophys. J. Lett.* **354**, L37–L40 (1990).
37. Hogan, C.J. *Nature* **344**, 107 (1990).
38. Hoyle, F. & Wickramasinghe, N.C. *Astrophys. Space Sci.* **147**, 245–256 (1988).
39. Dittman, W. & Neumann, K. in *Growth and Perfection in Crystals* (eds Daramus, R.N., Roberts, B.W., & Turnbull, D.) (Wiley, New York, 1958).
40. Seward, F.D., Hardeon, F.R. & Elsner, R.F. in *The Crab Nebula and Related Supernova Remnants* (eds Kafatos, M.C. & Henry, R.C.B.) (Cambridge University Press, 1985).
41. Helou, G., Soifer, B.T. & Rowan-Robinson, M. *Astrophys. J. Lett.* **298**, L7 (1985).
42. Burbidge, E.M., Burbidge, G.R. & Hoyle, F. *Astrophys. J.* **138**, 873–888 (1963).
43. Sandage, A.R. *Astrophys. J.* **138**, 863–872 (1963).
44. Burbidge, G. *Pub. Astr. Soc. Pacific* **98**, 1252–1257 (1986).
45. Burbidge, G. & Hewitt, A. *Comets to Cosmology* (ed. Lawrence, A.) 320–325 (Springer, Berlin, 1988).
46. Sandage, A.R. & Cacciari, C. *Astrophys. J.* **350**, 645–661 (1990).
47. Cowan, J.J., Thielmann, F.-K. & Truran, J.W. *Astrophys. J.* **323**, 543–552 (1987).
48. Das Gupta, P., Narlikar, J.V. & Burbidge, G. *Astron. J.* **95**, 5–14 (1988).
49. Das Gupta, P. thesis, Bombay Univ., 1988.
50. Schneider, D.P., Schmidt, M. & Gunn, J.E. *Astron. J.* **98**, 1951–1958 (1989).
51. Arp, H.C. & Burbidge, G.R. *Astrophys. J.* **353**, L1–L2 (1990).
52. Hoyle, F. & Narlikar, J.V. *Action at a Distance in Physics and Cosmology* (Freeman, San Francisco, 1974).
53. Boesgaard, A.M. & Steigman, G. *A. Rev. Astron. Astrophys.* **23**, 319–379 (1985).
54. Playfair, J. *Illustrations of the Huttonian Theory of the Earth* (Edinburgh, 1802).
55. Arp, H. *Astrophys. Lett.* **7**, 221 (1971).

# A highly mobile Laurentide ice sheet revealed by satellite images of glacial lineations

G. S. Boulton & C. D. Clark\*

Department of Geology and Geophysics, Grant Institute, Edinburgh University, West Mains Road, Edinburgh EH9 3JW, UK

\* Present address: Department of Geography, University of Sheffield, Sheffield S10 2TN, UK

Satellite images of the North American continent reveal previously unrecognized patterns of crossing drift lineations reflecting unexpected shifts of centres of mass of the Laurentide ice sheet throughout the last glacial cycle. Such shifts must have been associated with major global atmospheric circulation changes.

LATE Quaternary climate change was dominated by the growth and decay of large ice sheets in mid-latitudes, in particular of the North American (Laurentide) ice sheet. This ice sheet has been the dominant influence on global sea level change and played an important part in modulating the response of the global climate system to external forcing<sup>1-3</sup>. Although there is evidence that there were strong fluctuations of its margins during the last climate cycle, and although there is disagreement about whether the ice sheet was an elongate dome trending NW-SE over Hudson Bay<sup>4-7</sup> or whether it comprised two domes, one over Keewatin and one over Labrador and Quebec<sup>8-10</sup>, it has been generally concluded<sup>11,12</sup> that the centres of mass and flow were stationary throughout much of the glacial cycle.

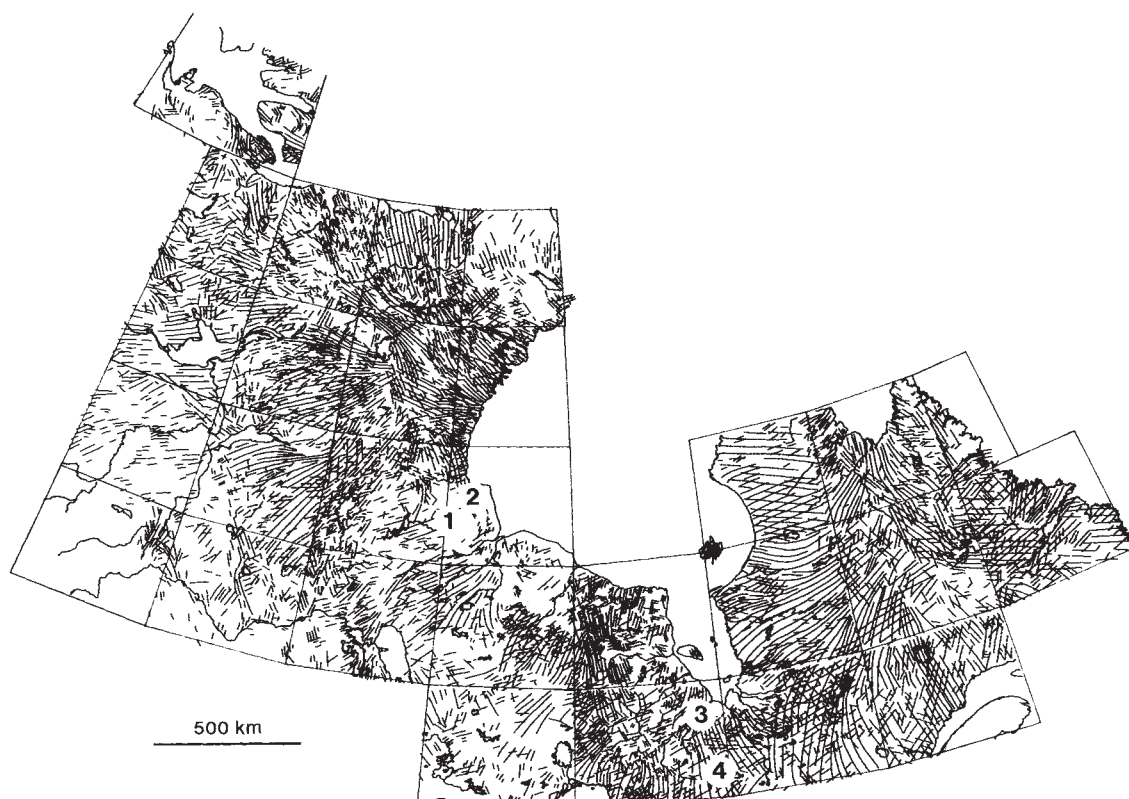
Our studies indicate much more dynamic ice-sheet behaviour, with strong fluctuations of the margins, particularly on the western flank, and shifts in the location of centres of mass and flow by up to 1,000-2,000 km.

An ice sheet with a static centre of mass but fluctuating margins may simply reflect a direct response to externally forced atmospheric change. Strongly shifting centres of mass, however, probably reflect strong atmosphere/ice sheet coupling, in which the changing atmospheric perturbation generated by a growing ice sheet affects the distribution of accumulation and ablation on its surface and thus the evolution of both the ice sheet and atmosphere in response to external forcing. This underlines the importance of coupled, time-dependent, three-dimensional ice sheet/atmosphere models.

## Crossing drift lineations

Satellite images show large-scale lineations, which aerial photography and ground surveys reveal to be sediment ridges oriented parallel to former glacier flow<sup>13,14</sup>. These are good indicators of integrated large-scale flow patterns as individual ridges are long and overlap with many others, defining continuous flow patterns over many hundreds of kilometres. Sets of lineations do cross, however, suggesting temporal changes in ice-sheet flow. We have mapped these lineations over large parts of mainland Canada (Fig. 1). Their relative ages can be established from high-resolution satellite images or conventional aerial photographs by the clear reorientation of elements of pre-existing lineations into new lineation sets<sup>15</sup>. Objective relative chronologies have been established for Keewatin, to the west of Hudson Bay, and the James Bay-Labrador-Quebec area, to its east (Fig. 2a).

FIG. 1 Trends of superimposed sets of drift lineations observed on Landsat images of mainland Canada. (This is a representative selection of a much greater lineation density.) The lineations are presumed to have formed parallel to the flow of an ice sheet with shifting centres of flow radiation. Numbers show the location of sections in Fig. 5.





### Survival of old lineations

In an ice sheet that flows over a bed of low relief, the direction of basal flow will be parallel to the surface slope. As the Laurentide was such an ice sheet we assume that the lineations formed parallel to basal flow during periods of sustained steady flow and suggest that the successive superimposed patterns of lineation reflect shifts in the centres of ice sheet mass and the associated patterns of flow.

Because the velocities of land-based ice sheets along flow lines increase from zero at ice divides to a maximum near to the equilibrium line (which separates the zones of accumulation and ablation) and then diminish towards the terminus, it has been argued<sup>12</sup> that erosion will be slight in the region of the ice divide but will increase to a maximum where the rate of velocity increase is greatest. Near the equilibrium line, velocities decrease and deposition dominates from there to the terminus (Fig. 3a). For a period of given duration the longest and strongest lineations will tend to occur beneath the outer part of the glacier, where erosion and deposition are strongest, but where earlier lineations will tend to be obliterated. Earlier lineations will tend to survive when shifting ice-sheet geometry superimposes a divide area over an earlier zone of rapid flow. A single, integrated radial lineation set may reflect a long period of steady state, or a growing or decaying ice sheet with a stable flow pattern. Distinctive crossing sets imply either relatively long periods of stable flow geometry with intervening phases of relatively rapid divide movement, or changes in glaciological regime. Relatively small shifts in the location of divides may produce large angular changes in lineation direction in the divide region (Fig. 3b, c). Angles between crossing lineations will tend to be least near ice-sheet margins.

### Relative chronology of lineation sets

We have compared the flow sets listed in Fig. 2a that do not cross, and are not therefore of demonstrably different age, to investigate whether they are compatible with a single flow pattern. For example, we find that although flow sets 13 and 3 (see Fig. 4) cannot be distinguished on the basis of cross-cutting relationships, a synchronous flow pattern could not have pro-

duced both. The following combinations do not show cross-cutting relations and are compatible with a single flow pattern: 1-21-23, 38-39-5-35-24-16, 34-7-4. Whereas 1-21-23 are demonstrably of the same age (they are the lineations formed during the final retreat<sup>4</sup>), 38-39-5-35-24-16 (and 34-7-4) could represent synchronous flows and are mutually compatible, but their synchronicity has not yet been demonstrated.

Using these three integrated flow sets as a framework, a long series has been established for Keewatin. Although there is no objective evidence for the relative age of flow sets 34-7-4 and 13, lineations in set 13 appear more degraded than those in 34-7-4, and we therefore tentatively assume that they are older. Sets 9 and 11 post-date 38-39-5-35-24-16, but their relationship to 1-21-23 is unknown. As the latter reflects the final pattern of flow during the decay of the last Laurentide ice sheet, however, 9 and 11 must pre-date it. In the James Bay lowlands and Labrador-Quebec, sets 22 and 20 (not illustrated) are so nearly congruent that we assume them to be roughly synchronous. They pre-date 18 and 38-39-5-35-24-16, but there is no evidence of direct correlation with sets in Keewatin.

We use this temporal sequence of sets, established partly on an objective basis and partly on the basis of simplicity, to suggest a sequence of flow stages A-G (Figs 2b, 4) that provide a framework in which stratigraphic information from widely separated sites can be integrated to develop a more complete history of ice-sheet behaviour.

### Correlations with stratigraphy and dating

There are many published sections south of Hudson Bay where fabric orientations in tills coincide with the orientations of the local flow sets. For instance, sections along the Churchill River<sup>16</sup> contain three tills (Fig. 5a, b) and fabrics in the lower two coincide with the local lineations of local flow set 39 (stage B) and flow set 34 (stage C). Flow set 16 of stage C is persistent

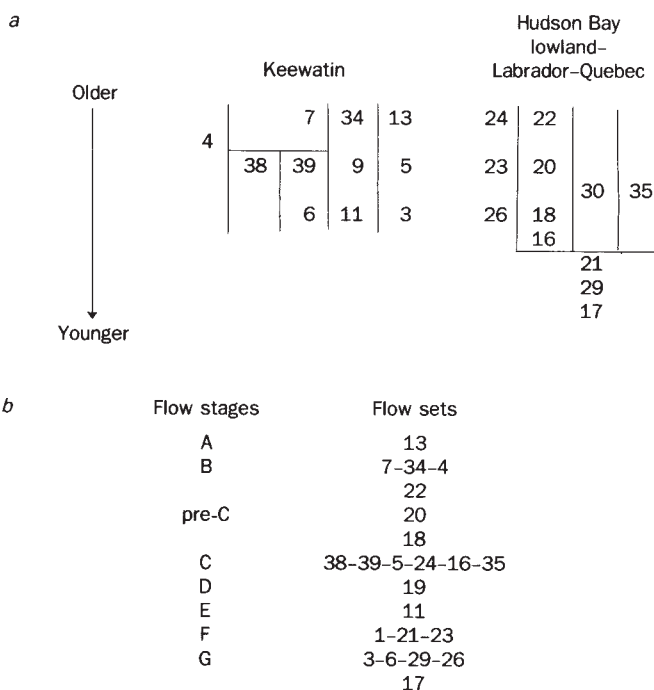


FIG. 2 a, Relative ages based on cross-cutting flow sets (vertical lines separate flow sets for which relative ages have not been established). b, Flow stages based on compatible flow sets. Figure 4 shows the location of the flow stages and flow sets.

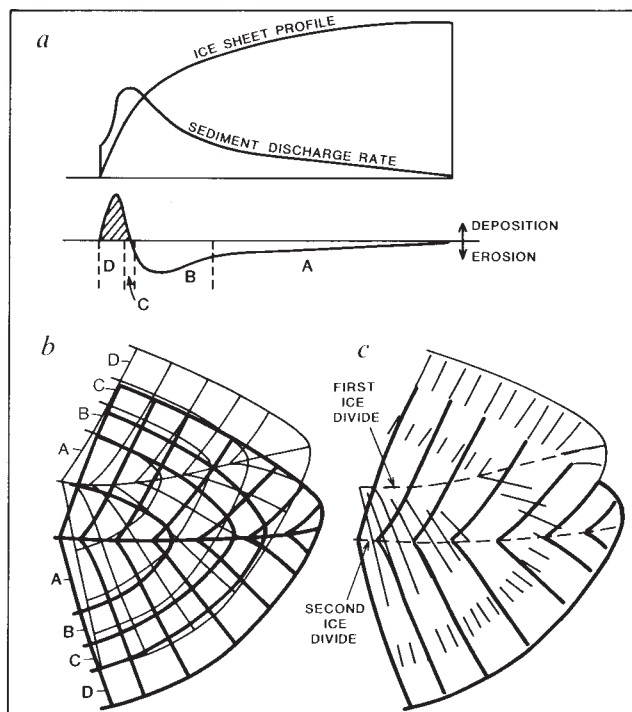


FIG. 3 a, Theoretical distribution of erosion and deposition beneath an ice sheet<sup>15</sup>. Zones of low erosion or deposition rate (A and C) are most likely to preserve earlier lineations. b, Changing pattern of ice flow (radial lines) and surface form (concentric lines) in part of an ice sheet as a consequence of ice-divide migration. Thin lines show pre-migration and thick lines post-migration features. c, Angles between crossing lineations are large in the zone of shifting ice divides. Post-shift locations of zones A and C favour preservation of earlier lineations.

throughout the area south of James Bay. It is reflected in fabrics from Skinner's<sup>17</sup> Adam Till in the Moose River basin (Fig. 5c), in the uppermost of the 'older tills' in the Timmins area<sup>18</sup>, in the earliest of three superimposed bedrock striation sets in the Abitibi-Timiskaming region<sup>19</sup>, and in the middle of a three-till sequence in the Matheson area<sup>20</sup> (Fig. 5d). In this latter area, the Middle Till with a fabric trend of 240° separates the Lower Till with an approximately N-S fabric trend from the Upper Till (the Matheson Till) with a similar trend (Fig. 5). From these relationships and fabric orientations in the tills, we suggest that these three units represent flow stages B, C and F. Correlation with the aminostratigraphy of the Hudson Bay lowlands<sup>21,22</sup> (Fig. 5) suggests that stage B occurred between 106 and 76 kyr and that stage F occurred between 35 and 8 kyr.

We do not expect the full sequence of flow stages to be everywhere represented in the till sequence because the drift lineations that we have mapped must represent superimposed lenticles of till that are laterally impersistent. We have therefore only used correlation sequences that seem to show a fuller record.

Shilts<sup>21</sup> and Andrews *et al.*<sup>22</sup> proposed that marine inundation occurred in the Hudson Bay lowlands at 106 kyr, 76 kyr and 35 kyr before present and that this is more likely to have occurred if the Laurentide ice sheet, rather than having a dome over Hudson Bay, had domes over Keewatin and Labrador-Quebec, as it did during the final stage of decay, and that there was a thin zone of confluence in Hudson Bay which was easily broken up, allowing marine penetration along a narrow corridor at 35 kyr and 76 kyr.

Shilts<sup>8</sup> has argued that the general distribution of erratics supports the two-dome concept. Palaeozoic carbonate bedrock underlies much of Hudson Bay and was glacially dispersed towards the south and southeast, but not towards the west or east, whereas distinctive red erratics from the Dubawnt Group in Keewatin are dispersed towards the south, east and northeast. He proposed a pattern of contemporaneous flow from confluent Keewatin and Labrador-Quebec domes that would generate the observed pattern of dispersal. Our evidence suggests, however, that the centres of outflow of the Laurentide ice sheet were not static, as implied by Shilts, but highly mobile, and that the pattern of dispersal could reflect changing rather than stable patterns of flow.

### The ice sheet through the glacial cycle

Correlation with the Hudson Bay lowlands sequence and acceptance of the amino-acid chronology of Andrews *et al.*<sup>22</sup> suggests that stages C–G, at least, belong to the last (Wisconsinan) glacial cycle. Below we summarize the principal flow stages and the extent to which they are compatible with other geological data (Fig. 4).

**Flow stage A.** (Set 13) This flow is derived from the extreme northeast of Keewatin, from the direction of the Gulf of Boothia, Melville Peninsula and northwest Baffin Island. It is consistent with 'early striations' observed by Tyrell<sup>23</sup> and with evidence of dispersal trains directed from the north in the vicinity of Baker Lake<sup>24</sup>. Although we have no direct evidence of its age, it may reflect earliest Wisconsinan ice-sheet expansion from centres of initiation in northern Baffin Island or northeast Keewatin, because cooling could lead to ice-sheet initiation in Keewatin as readily as in Labrador<sup>25</sup>.

**Flow stage B.** (Set 34–7–4) There is no evidence for the age of this stage. It may correlate with either or both of the lowest tills at Mountain Rapids and Limestone Rapids (Fig. 5) although Dredge and Nielsen<sup>16</sup> imply that the former may be of pre-Wisconsinan age. The stage reflects expansion of ice into Manitoba and Saskatchewan and is associated with a south-westerly movement of the ice divide of at least 400 km.

**Pre-C stages around James Bay.** (Sets 22–20, 18, not illustrated) Pre-C stages have been identified but not directly related to stages in Keewatin. Sets 22–20 seem to reflect a dominant flow

from Quebec into the James Bay lowlands. They are succeeded by set 18 which reflects flow into the same area from the north-west, presumably from Keewatin. The sequence may reflect earliest Wisconsinan development of a Labrador-Quebec dome, followed by the dominance of a Keewatin dome, before both coalesced to form the Hudson Bay ridge of stage C. They may, however, be of pre-Wisconsinan age.

**Flow stage C.** (Sets 38–39–5–35–24–16) Stratigraphic correlations in the Hudson Bay lowlands (Fig. 5) indicate that this stage is of early Wisconsinan age, between 106 kyr and 76 kyr.

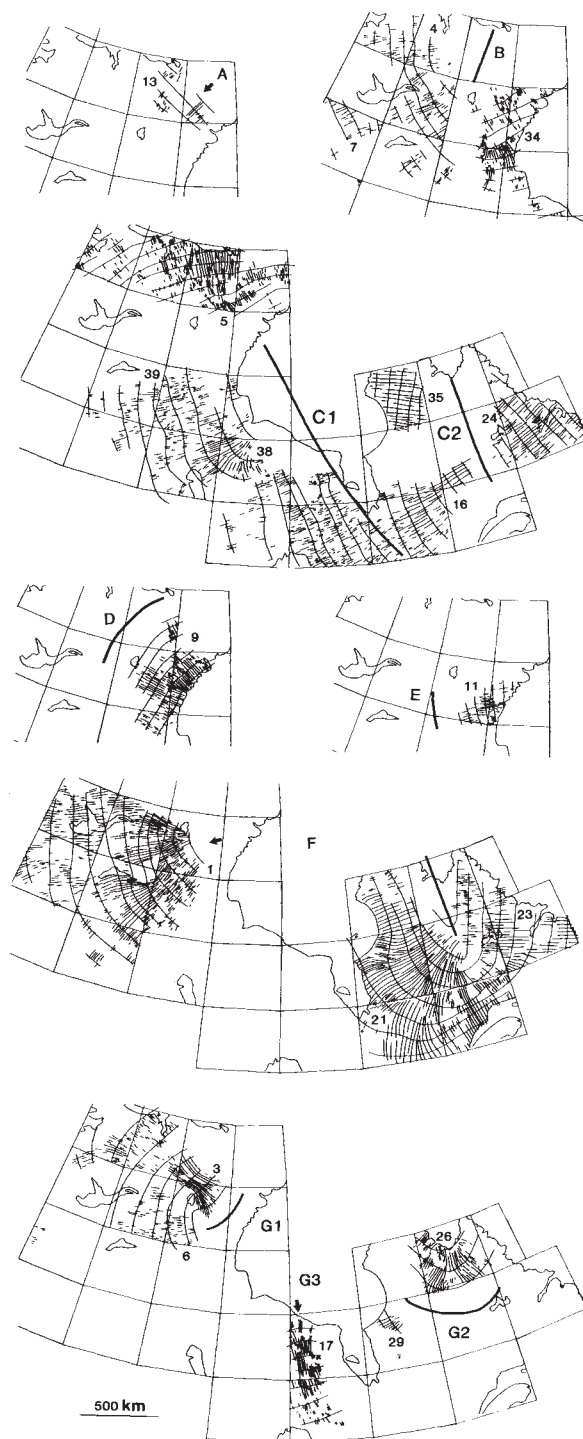


FIG. 4 The inferred sequence of flow stages (A–G). Locations of flow centres or divides are shown by thick lines. C1 shows an early Hudson Bay ice ridge and C2 a later, residual Labrador-Quebec ridge. The numbers label each ice-flow set.



It cannot, however, represent a single synchronous pattern of flow because normals to flow lines define an ice centre in Labrador and it is inconceivable that, on an approximately horizontal bed, ice from a Labrador source should penetrate into Alberta some 2,500 km away, but that its extension to the east should be limited by the continental shelf edge, a distance of 1,100 km. Instead, we suggest that the lineation set reflects progressive decay of a large ice sheet that near its maximum had an ice divide orientated northwest-southeast over central Hudson Bay, and which progressively shifted eastwards as the western part of the ice sheet decayed. The slowly retreating glacier continuously reorganized its bed in the geomorphologically active terminal zone producing colinear features radial to a final ice centre in Labrador. The conditions necessary to maintain the ice sheet deteriorated in the west but were sustained in the east.

We suggest that the early stage C Laurentide ice sheet is equivalent to the early Wisconsinan ice sheet. It is considered by many to have been more extensive than the late Wisconsinan ice sheet in eastern Canada, where it extended far out onto the continental shelf<sup>26</sup>, in its southwestern sector in the northern United States<sup>27</sup> and on its northern margins in east Baffin Island and Banks Island<sup>27-29</sup>.

The deglaciation of Hudson Bay late in stage C was probably contemporaneous with a period of marine inundation at ~75 kyr<sup>22</sup> (Fig. 5) and the known reduction in global ice volume. Late stage C (and stage D) could be of middle Wisconsinan age, during which a non-glacial period is recognized in western Canada<sup>30</sup>, and sites in southern Manitoba, central Saskatchewan and Alberta indicate an ice-free period from at least 40 kyr to 22 kyr<sup>31</sup>.

Of the stages we have identified, this is likely to be the one during which limestone erratics from Hudson Bay underwent transport towards the southwest. The northern limit of this dispersal zone on the western side of Hudson Bay is an almost precise continuation of the edge of the Palaeozoic limestone outcrop in Hudson Bay<sup>32</sup> and almost parallel to the flow lines of stage C. A slight deflection of ice flow in the Bay against its western shore could have ensured that the flow was parallel to the edge of limestone outcrop.

**Stage D.** (Set 9) This stage indicates easterly ice flow into Hudson Bay and must reflect redevelopment of a dome in Keewatin. The convergence of flow towards Hudson Bay is most probably explained by rapid flow into a water body (marine drawdown) and suggests that marine conditions persisted in Hudson Bay after late stage C. There is no clear indication of age for stage D.

**Stage E.** (Set 11) This shows northeasterly ice flow into Hudson

Bay from an ice centre that has shifted to the southwest. There is no direct evidence for its age.

**Stage F.** (Sets 1-21-23) The bi-nodal radial patterns shown by the lineation sets of this stage are very similar to the patterns of flow associated with the reconstructed decay of the late Wisconsinan Laurentide ice sheet<sup>4,9,33</sup>. We therefore conclude that stage F is of late Wisconsinan age. Shilts<sup>8</sup> has argued that even at its maximum extent, the late Wisconsinan ice sheet had domes in Keewatin and Labrador-Quebec, whereas others<sup>5,11</sup> have argued for a single ridge over Hudson Bay. Although the stage F lineations in Keewatin are compatible with either a Keewatin Dome, or a Hudson Bay ridge terminating in Keewatin, the stage F flow pattern in Labrador-Quebec indicates ice flow directly west into Hudson Bay thus precluding a trans-Laurentide ice ridge. Existence of a dome over Labrador requires that the Keewatin flow set reflects a dome, the two being connected by a saddle. This situation could have succeeded a ridge across Hudson Bay connecting Labrador with Quebec<sup>4</sup>, although we would have expected some reflection of this in the pre-F lineation pattern on the east side of Hudson Bay.

We suggest that the Keewatin dome was the successor to the growing stage D and E ice masses in Keewatin, and that the Labrador-Quebec dome grew from the residual stage C dome in that area, which many believe could have persisted there from early Wisconsinan times<sup>34,35</sup>.

**Stage G.** (Set 3-6, 17 and 26-29) There are three post-F sets that we are unable to relate to each other, but for convenience have grouped together.

There seems to have been a minor shift in the flow pattern on the western side of the Keewatin dome, represented by sets 3-6, related to a southeastern shift in the ice divide. We associate this with stage G1.

In stage F in Labrador-Quebec there seems to have been an ice ridge oriented north-south over Ungava Bay which may have led to erratic transport from Labrador to southern Baffin Island across the Hudson Strait<sup>36</sup>. Lineation set 26 of stage G2 suggests, however, that an embayment later developed in the Labrador-Quebec ice dome in the Ungava Bay hinterland, presumably as a result of marine drawdown. This would have permitted the Hudson Strait to act as a thoroughway for ice flowing from the Laurentide heartland to the Labrador Sea.

This reconstruction is compatible with the dispersal train of erratics from the Dubawnt group, which trend in a southeasterly direction from their source for 400-500 km before turning abruptly to the northeast<sup>8</sup>. The early flow to the southeast could reflect flow during the growth of the Keewatin dome before the development of a Hudson Strait ice stream, and the northeasterly

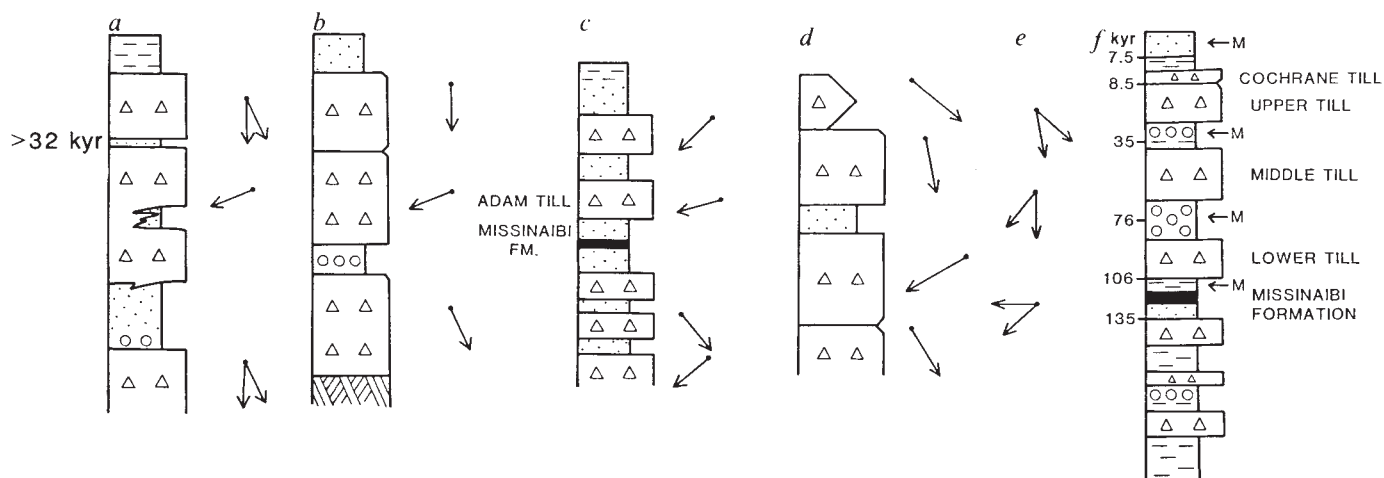


FIG. 5 Some sections for which till fabric data allow flow stages to be correlated with stratigraphy. Arrows show fabric data. *a*, Mountain Rapids, Churchill River<sup>16</sup>. *b*, Limestone Rapids, Churchill River<sup>16</sup>. *c*, Moose River

basin<sup>17</sup>. *d*, Matheson area<sup>20</sup>. *e*, Abitibi-Tamiskaming<sup>19</sup>. *f*, Summary<sup>22</sup> of Hudson Bay lowlands.

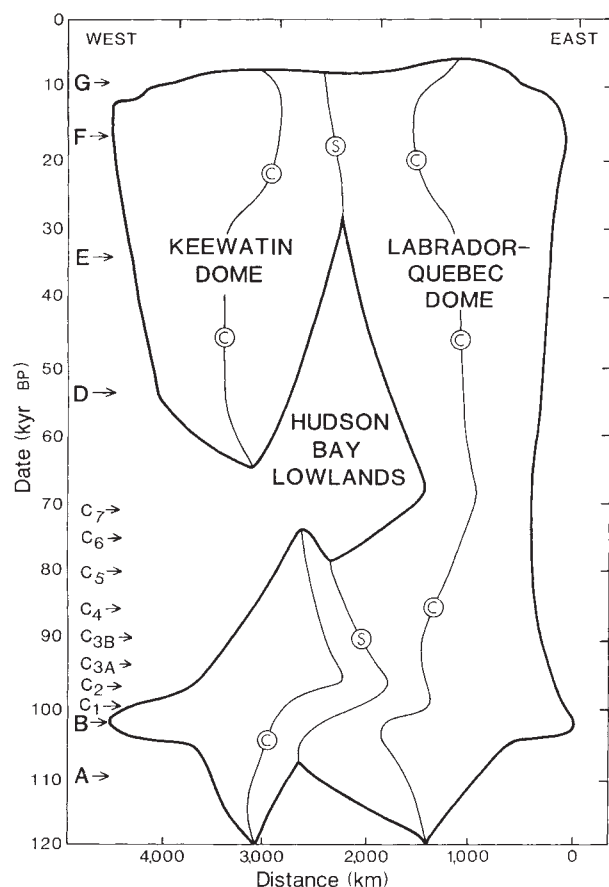


FIG. 6 Postulated time/distance envelope of ice sheet extent of an east-west section through Canada through the last glacial cycle. C shows inferred ice-sheet culminations (domes or ridges) and S saddles between domes or ridges.

shift may reflect the progressive drawdown associated with the development of flow through the Hudson Strait. The F-G1 change in the Keewatin ice centre might also reflect this change.

G3 lineations (set 17) are very strong in the area west of James Bay. We presume that they were formed during the co-called Cochrane surge just before 8 kyr<sup>4</sup>.

### Time-distance envelope of the ice sheet

Figure 6 shows an east-west time-distance transect across Hudson Bay. It is a schematic representation of the suggested move-

ment of Laurentide ice-sheet margins, divides and saddles during the Wisconsin glacial stage. Ice-sheet margins are smoothed. Stage C and F margins are assumed to be the early and late Wisconsin margins reconstructed by Vincent and Prest<sup>37</sup>, with other margins estimated from our analysis of flow stages. Complete flow patterns at any one stage are extrapolated from the partial geological data. Inter-stage flows and ice margins are interpolated.

We find only one long ice-free period in early to mid Wisconsin times in the area of the James Bay lowlands, whereas Andrews *et al.*<sup>22</sup> argue for two. This reflects our failure to find a lineation set that could be correlated with a mid-Wisconsin glacial event, which might have subdivided this period.

### Conclusion

The preservation of old lineations beneath more recent ones is an unexpected result, and provides a tool whereby palaeo-ice flow patterns can be reconstructed, and the magnitude of long-term glacier erosion and evolving patterns of erosion and deposition can be assessed. The reconstruction of integrated flow patterns in the interior of an evolving ice sheet is also a powerful means of using stratigraphic and chronological information from widely spaced localities. By combining these data we find that the Laurentide ice sheet's behaviour was much more dynamic than the presumed long-term behaviour of modern ice sheets and previous views of Pleistocene mid-latitude ice sheets.

Broccoli and Manabe<sup>3</sup> have demonstrated the topographic effect of the Laurentide ice sheet in perturbing the large-scale atmospheric flow over the ice sheet, generating winds from a southerly sector in the west, and a northerly sector in the east. These contrasts would clearly affect the mass balance on the ice-sheet surface and imply very strong coupling between ice sheet and atmosphere. Analysis of the significance of the inferred changes for the operation of the climate system clearly requires a three-dimensional, time-dependent, coupled ice sheet/atmosphere model. The ice-sheet history that we propose must have been associated with important downstream effects on the atmosphere and ocean in the Atlantic sector. The role of crustal isostasy in modulating ice-sheet response to external forcing<sup>1,38</sup> may also have been important in influencing the events that we have reconstructed.

We are aware of the potential errors in amino-acid chronologies<sup>39</sup> such as that of the Hudson Bay lowlands, and that the sequence of flow stages A-G may therefore reflect more than one glaciation. If this is the case, the results would be no less interesting, and could be evidence of non-unique, possibly 'chaotic' processes in ice-sheet initiation and evolution. This possibility underlines the importance of precise dating of the Hudson Bay lowlands sequence. □

Received 29 March; accepted 11 July 1990.

- Peltier, W. R. *A. Rev. Earth planet. Sci.* **9**, 199-226 (1981).
- Broecker, W. S. *et al. Nature* **341**, 318-321 (1989).
- Broccoli, H. J. & Manabe, S. *Géogr. phys. Quat.* **41**, 291-299 (1987).
- Dyke, A. S. & Prest, V. K. *Géogr. phys. Quat.* **41**, 237-263 (1987).
- Denton, G. H. & Hughes, T. J. (eds) *The Last Great Ice Sheets* (Wiley, New York, 1981).
- Sugden, D. E. *Arct. Alp. Res.* **9**, 21-47 (1977).
- Peltier, W. R. *A. Rev. Earth planet. Sci.* **9**, 199-226 (1981).
- Shilts, W. W. *Nature* **286**, 213-218 (1980).
- Boulton, G. S., Smith, G. D., Jones, A. S. & Newsome, J. *J. geol. Soc. Lond.* **142**, 447-474 (1985).
- Fisher, D. A., Reeh, N. & Langley, K. *Géogr. phys. Quat.* **39**, 229-238 (1985).
- Flint, R. F. *Geol. Soc. Am. Bull.* **54**, 325-362 (1943).
- Aylsworth, J. M. & Shilts, W. W. *Geol. Surv. Can. Pap.* **88-24** (1989).
- Boulton, G. S. & Clark, C. D. *Trans. R. Soc. Edin.* (in the press).
- Clark, C. D. *IEEE Trans. Geosci. Remote Sensing* (in the press).
- Boulton, G. S. in *Drumlin Symp.* (eds Menzies, J. & Rose, J.) (Balkema, Rotterdam, 1987).
- Dredge, L. A. & Nielsen, E. *Geol. Surv. Can. Pap.* **85-1A**, 247-257 (1985).
- Skinner, R. G. *Geol. Surv. Can. Bull.* **225** (1973).
- Dilabio, R. W., Miller, R. F., Mott, R. J. & Coker, W. B. *Geol. Surv. Can. Pap.* **88-1C**, 61-65 (1988).
- Veilleux, J. J. *Can. J. Earth Sci.* **23**, 1724-1742 (1986).
- Steele, K. G., Baker, C. L. & McClenaghan, M. B. *Geol. Surv. Can. Pap.* **89-20**, 127-138 (1989).

- Shilts, W. W. *Naturaliste Can.* **109**, 309-332 (1982).
- Andrews, J. T., Shilts, W. W. & Miller, G. *Quat. Res.* **19**, 18-37 (1983).
- Tyrell, J. B. *J. Geol.* **6**, 147-160 (1988).
- Cunningham, C. M. & Shilts, W. *Geol. Surv. Can. Pap.* **64-20** (1977).
- Williams, L. D. *Arct. Alp. Res.* **11**, 443-456 (1979).
- Grant, D. R. *Géogr. phys. Quat.* **31**, 247-260 (1977).
- Clayton, L. & Moran, S. R. *Quat. Sci. Rev.* **1**, 55-82 (1982).
- Nelson, A. R. *Geol. Soc. Am. Bull.* **92**, 512-518 (1981).
- Vincent, J. T. *Geol. Soc. Am. Pap.* K-3, 97-104 (1987).
- Fulton, R. J., Fenton, M. M. & Rutter, N. W. *Geol. Surv. Can. Pap.* **84-10**, 69-86 (1984).
- Dredge, L. A. & Thorleifson, L. H. *Géogr. Phys. Quat.* **41**, 215-235 (1987).
- Dredge, L. A. *Can. J. Earth Sci.* **25**, 783-787 (1988).
- Prest, V. K. *Geol. Surv. Can. Map* 1257A (1969).
- King, L. H. & Fader, G. B. *J. Geol. Surv. Can. Bull.* **363** (1986).
- Piper, D. J. W. *Geosci. Can.* **15**, 23-28 (1988).
- Andrews, J. T., Stravers, J. A. & Miller, G. H. in *Models in Geomorphology* (ed. Walden, M.) 93-117 (Allen & Unwin, London, 1985).
- Vincent, J. S. & Prest, V. K. *Géogr. Phys. Quat.* **41**, 199-213 (1987).
- Pollard, D. *Nature* **272**, 233-235 (1978).
- Brigham-Grette, J. & Miller, G. H. *Int. Un. Quat. Res. 12th Congr. Prog. and Abstr. Vol. 136* (Natl. Res. Council, Ottawa, 1987).



# *In vitro* selection of RNA molecules that bind specific ligands

Andrew D. Ellington & Jack W. Szostak\*

Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

Subpopulations of RNA molecules that bind specifically to a variety of organic dyes have been isolated from a population of random sequence RNA molecules. Roughly one in  $10^{10}$  random sequence RNA molecules folds in such a way as to create a specific binding site for small ligands.

THE probability that a given random sequence polynucleotide or peptide chain will fold to form a stable three-dimensional structure with a given ligand binding or catalytic activity is unknown, but is generally thought to be very low. A related problem, which has been equally difficult to address theoretically, is to estimate the number of fundamentally different classes of structures capable of carrying out a given binding or catalytic function (for example, hundreds of different proteases have been identified, but these fall into only five independent structural classes). These questions are important for theories of the origin and early evolution of life, as the first biological catalysts presumably arose from random sequence polymers<sup>1-3</sup>. We have started to address these questions by developing methods for the synthesis of large numbers of random sequence RNA molecules, and for the isolation of molecules with specific ligand binding properties from such populations.

## Synthesis of random sequence RNAs

We began by chemically synthesizing a large pool of DNA molecules made up of an average of 100 bases of random sequence flanked by defined regions at the 5' and 3' ends. The defined sequences were necessary for primer hybridization, and also contained restriction sites to facilitate cloning. The 155-nucleotide (nt) long DNAs were synthesized by solid-phase phosphoramidite chemistry on a Biosearch 8700 automated synthesizer, using an equimolar mixture of the four bases for the random portion of the sequence. On the basis of the yield of synthesized oligonucleotides (100 µg), the complexity of this pool corresponds to approximately  $10^{15}$  individual sequences.

Preliminary experiments indicated that these synthetic DNA molecules were not efficiently transcribed into RNA by T7 RNA polymerase, perhaps because of incomplete deprotection. To obtain a larger amount of fully active DNA suitable for repeated use, the initial 'random-mers' were amplified 20-fold by large-scale polymerase chain reaction (PCR) to generate a sequence library. Transcription of this pool allowed the reproducible synthesis of high yields of RNA. The quality of the pool was evaluated by sequencing 14 clones derived from the initial pool of DNA, and an additional 14 clones derived from the PCR-amplified DNA. The base composition is roughly random, with a slight excess of G. There is no bias in the distribution of di- and tri-nucleotides (data not shown). As expected from a pool of such large complexity, all 28 sequences were different, and showed no similarity to each other.

A considerable decrease in pool complexity is likely to have occurred during the PCR amplification because of inability to replicate past lesions on the chemically synthesized DNA and because of poor replication of a very GC-rich portion of the 5'

defined sequence. Primer extension experiments have shown that only 4–5% of the synthetic DNA template can be completely copied, leading to a 20–25-fold decrease in pool size. In addition, sequencing showed that about two thirds of the amplified molecules had mutations in the GC-rich region after PCR; if these were preferentially amplified from rare mutants, the pool complexity could have been reduced by an additional factor of three or more. We estimate that the effective complexity of the final amplified pool corresponds to  $\sim 10^{13}$  different sequences.

## *In vitro* selection

The double-stranded DNA molecules generated by PCR have a T7 RNA polymerase promoter at one end; *in vitro* transcription with T7 RNA polymerase results in the formation of a pool of RNA random-mers. This RNA pool was purified by denaturing polyacrylamide gel electrophoresis to ensure size homogeneity, then subjected to a brief incubation at 70 °C to allow conformational equilibration. RNA molecules in this population that were capable of binding to specific ligands were then enriched by affinity column chromatography. About 30 µg of this RNA ( $\sim 4 \times 10^{14}$  molecules, or an average of 40 copies of each of the  $10^{13}$  different sequences in the pool) was applied to the affinity column in a high-salt buffer (0.5 M LiCl) to allow the negatively charged RNAs to approach negatively charged ligands. Non-binding species were washed off the column with three column volumes of high salt buffer, and specifically bound material was then eluted with water.

Eluted RNAs were treated with reverse transcriptase to generate complementary DNAs, which were then amplified by PCR (Fig. 1). In the first cycle, only one third of the eluted RNA was converted to cDNA. Transcription of the resulting double-stranded DNA regenerates RNA, which can then be used for another cycle of enrichment. Multiple cycles of affinity chromatography and *in vitro* amplification result in the continued purification of binding species, until virtually the entire population is capable of binding to the column.

As an initial test of this method, we selected RNAs that would bind to an oligo(dT)-Sepharose (OT) column. Less than 1% of the applied RNA bound to the column in the first two cycles, but on the fourth cycle over 50% of the applied RNA bound to the oligo(dT) column (Table 1). We cloned and sequenced seven oligo(dT)-binding RNAs and, as expected, all of these contained oligo(A) stretches of from 10–24 nucleotides in length.

## Selection on dye columns

We then used the *in vitro* selection system to isolate RNAs that bind to several dyes that appear to mimic metabolic cofactors. For example, Cibacron Blue (Fig. 2) binds tightly to the NAD-binding site of many dehydrogenases and Cibacron Blue columns have been used for purification of these proteins by affinity chromatography<sup>4</sup>. The experiments described here used Cibacron Blue 3GA (CB), Reactive Red 120 (R), Reactive Yellow 86 (Y), Reactive Brown 10 (BR), Reactive Green 19 (GR) and Reactive Blue 4 (B4) attached to cross-linked, beaded agarose. We chose these molecules because they have many possible hydrogen-bond donor and acceptor groups as well as planar surfaces for stacking interactions.

Each of the six organic dye columns initially bound less than 0.1% of the applied RNA, but after four to five cycles of selection

\* To whom correspondence should be addressed.

TABLE 1 Multiple selection cycles yield dye-binding aptamers

|                      |   | Dye column         |                    |                    |                    |                 |                 |                 |
|----------------------|---|--------------------|--------------------|--------------------|--------------------|-----------------|-----------------|-----------------|
|                      |   | CB                 | R                  | Y                  | BR                 | OT              | GR              | B4              |
| Cycle number         | 1 | 0.08               | 0.1                | 0.08               | 0.07               | 0.7             | 0.02            | 0.1             |
|                      | 2 | 2.0                | 0.2                | 1.0                | 0.4                | 0.8             | 0.9             | 0.4             |
|                      | 3 | 0.06               | 0.1                | 0.08               | 0.05               | 1.7             | 0.4             | 0.1             |
|                      | 4 | 1.6                | 3.7                | 2.1                | 2.7                | 55              | 65              | 32              |
|                      | 5 | 49                 | 42                 | 57                 | 69                 | 85              | 69              | 59              |
|                      | 6 | 63                 | 55                 | 57                 | 78                 | (82)            | (69)            | (61)            |
| Amplification factor |   | $2 \times 10^{10}$ | $4 \times 10^{10}$ | $2 \times 10^{10}$ | $5 \times 10^{10}$ | $2 \times 10^6$ | $3 \times 10^8$ | $1 \times 10^9$ |

The fraction of RNA which bound to a given column in each cycle (% RNA retained per cycle) was calculated by dividing the number of counts eluted by the number of counts originally loaded. The 'Amplification factor' is the inverse of the product of the fraction bound in each cycle. The sixth round values for OT, GR, and B4 (in parentheses) are replicates of the fifth round.

and amplification, every column bound over 50% of the applied RNA. The cumulative amplification factor after five cycles ranged from  $3 \times 10^8$  for GR to  $5 \times 10^{10}$  for BR. Assuming that all of the molecules capable of specific binding were in fact retained in each column cycle, the amplification factor for each cycle represents the degree of enrichment of the specific binding sequences. By dividing the estimated initial pool size by the cumulative amplification factor, we calculate that each of the dye-binding RNA pools is a complex mixture of  $10^2$ – $10^5$  different sequences. The apparent complexity of the selected pools implies that there are many independent sequence 'solutions' to each ligand-binding 'problem.'

### Binding of selected RNAs is specific

The binding specificity of the dye-selected RNA pools was tested by measuring the binding of each pool to each column. Three of the six pools were found to bind only to their cognate ligands (CB, GR and B4). The R pool bound preferentially to the R column but to a lesser extent to the other dye columns, and two of the pools (Y and BR) contained species that bound to all columns, either because they failed to discriminate between the different organic dyes or because they bound to the matrix itself (Table 2). The structural similarity between CB and B4 (Fig. 2) implies that the RNAs in these pools have binding sites capable of discriminating between functional groups on ligands.

We examined the binding properties of individual sequences from the CB and B4 pools by cloning PCR-amplified DNA from both pools. Individual RNA species were generated from each clone by using PCR to reintroduce the T7 promoter and primer binding sequences, followed by T7 transcription of the PCR DNA. We have termed these individual RNA sequences 'aptamers', from the Latin '*aptus*', to fit. RNAs from the cloned aptamers were examined for binding to their cognate ligands and exhibited a range of affinities (Table 3), as measured by column retention. Eight of the CB-binding aptamers were examined for binding to the Y column. Only one bound, thus seven out of eight were specific for CB binding.

We sequenced 17 clones from the CB pool, three of which were identical, and 14 clones from the B4 pool, all of which were different. We also sequenced 14 clones from the R pool and found two identical clones, while 10 clones from the GR pool were all different; this level of duplication suggests that the total complexity of these pools is in the range of 100–1,000 sequences. Most of the clones had no detectable similarity to any of the other clones, confirming the idea that there are many completely different ways of making specific binding sites for any given ligand. Careful sequence comparison did reveal short regions of similarity between several pairs of aptamers from each pool, suggesting that these regions might be involved in binding.

FIG. 1 Schematic diagram of the *in vitro* selection cycle. DNA random-mers (top) consist of 100 bases of random sequence flanked by defined regions. PCR-amplified DNA is transcribed into RNA, and RNA molecules of the desired binding properties are selected by affinity chromatography. Bound RNA is eluted, converted to cDNA, and amplified by PCR, at which point the cycle can be repeated.

**METHODS.** The initial RNA pools were synthesized by T7 RNA polymerase transcription of the PCR-amplified random-mers (500 µg DNA in a 5 ml reaction), subsequent RNA pools were synthesized by transcription of 1–2 µg of PCR DNA in an 80 µl reaction mix containing 3.2% PEG 8000, 5 mM each ATP, GTP, UTP and CTP, 5 mM DTT, 40 mM Tris-HCl, pH 7.9, 26 mM MgCl<sub>2</sub>, 0.01% triton X-100, 1 mM spermidine, 60 U RNasin (Promega), 300 U T7 RNA polymerase (New England Biolabs) and 20 µCi of [ $\alpha$ -<sup>32</sup>P]GTP. After 16 h at 37 °C, samples were treated with 2U of RQ1 DNase for 1 h at 37 °C to remove all DNA template. RNA transcripts were purified on 6% acrylamide, 7 M urea gels as previously described<sup>6</sup>. About 30 µg of RNA in 0.6 ml of 0.5 M LiCl, 20 mM Tris-HCl, pH 7.6, 1 mM MgCl<sub>2</sub> (column buffer) was heated to 70 °C for 5 min, allowed to cool to room temperature and applied to a 3.5 ml column of immobilized ligand (dyes linked to 4% cross-linked agarose, Sigma; oligo(dT)<sub>12–18</sub> on cellulose, Pharmacia). The column was then washed with 13 ml of column buffer and bound species were eluted with 10 ml of water. In cycles 4, 5 and 6, the water elution was followed by a 2 mM EDTA elution to remove tightly bound species. Eluted material was precipitated using 200 µg of glycogen as carrier, and 1/3 to 1/60 of the total eluted RNA was converted to cDNA as described<sup>7</sup> using an 18-mer primer complementary to the 3' end (5'-AAGCTTCCCGGGCTGCAG-3'). The cDNA was amplified to 3–6 µg of double-stranded DNA by PCR<sup>8</sup> with a 34-mer primer complementary to the 5' defined region; (5'-TAATACGACTCACTATAGGGAG-AATTCGCCGCGC-3'). A shorter version of this primer (a 29-mer) was used to amplify the initial DNA pool. PCR DNA was purified by chloroform extraction and a Sephadex-G50 spin column before transcription for subsequent cycles.

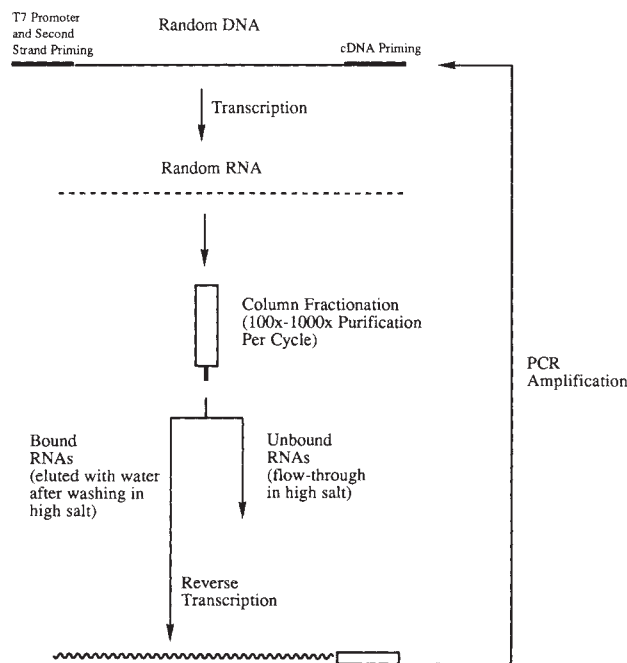




TABLE 2 Cross-binding of selected aptamer pools

|              |    | Dye column |    |    |    |    |    |    |
|--------------|----|------------|----|----|----|----|----|----|
|              |    | CB         | R  | Y  | BR | OT | GR | B4 |
| Aptamer pool | CB | 63         | 3  | 8  | 8  | 7  | 7  | 8  |
|              | R  | 16         | 79 | 18 | 15 | 12 | 16 | 15 |
|              | Y  | 76         | 32 | 62 | 66 | 4  | 66 | 50 |
|              | BR | 59         | 28 | 70 | 79 | 5  | 70 | 41 |
|              | OT | 6          | 5  | 7  | 7  | 61 | 6  | 5  |
|              | GR | 1          | 1  | 1  | 1  | 4  | 73 | 1  |
|              | B4 | 1          | 1  | 1  | 1  | 7  | 1  | 85 |

Values are % retained. After the sixth round of selection, the water-eluted RNA pools were diluted 1:1 with 2× column buffer and reappplied to each of the columns (thus, for the cognate columns this represents a seventh round of selection). Columns were run as described in Fig. 1, except that samples were eluted with 2 mM EDTA rather than water. The values reported are the per cent of total counts in the elution.

We chose two sequences for further analysis, CB-42 and B4-25. Each of these RNAs bound specifically only to the dye column they were selected on (data not shown). In addition, an analytical affinity column was run under isocratic conditions to determine an approximate dissociation constant ( $K_d$ ) for the RNA-ligand interaction. The  $K_d$  is roughly the concentration of ligand on the column multiplied by the ratio of the column volume (minus the void volume) to the elution volume (minus the void volume) (ref. 5). The B4 column contains about 10 mM bound ligand; the peak of B4-25 elution from a 10 ml column (with a 7 ml void volume) was at 54 ml, corresponding to a  $K_d$  of ~600  $\mu$ M; the CB column contained about 2 mM bound ligand, and from similar calculations CB-42 was found to have a  $K_d$  of less than 100  $\mu$ M. These dissociation constants are similar to those found for ligand binding of protein enzymes.

### Mapping binding sites

To show that individual aptamers contained well defined binding sites, we mapped the positions required for function by mutagenizing the entire aptamer sequence and then selecting for functional variants. We synthesized DNA oligonucleotides with the CB-42 and B4-25 sequences, such that each position was mutagenized at a level of 15% (that is, 85% correct nucleotide and 5% of each non-wild-type base). RNAs were made from

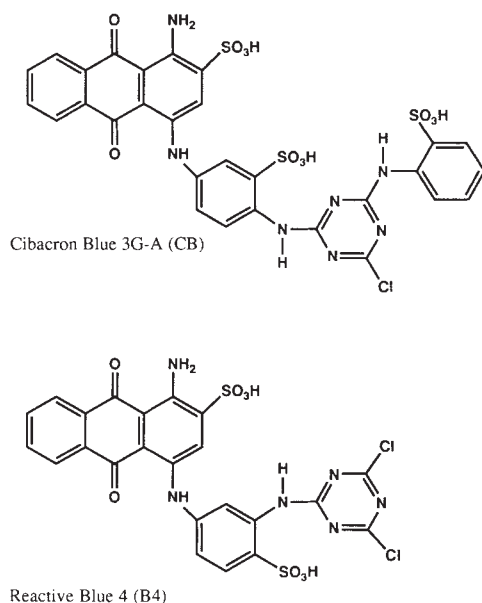


FIG. 2 Structures of two of the dyes used for affinity column selection of aptamers. Binding to the matrix is through the positions occupied by chlorine atoms in the free dyes.

these degenerate oligonucleotides as before, then subjected to cycles of *in vitro* selection and amplification. About 1% of applied RNA bound in the first cycle, but more than 70% of the sequences bound in the second. Assuming that the low binding-affinity of the initial pool is the result of mutations that destroy the binding site, we can estimate the number ( $n$ ) of nucleotides that must remain unchanged in order to retain binding activity from  $(0.85)^n = 0.01$  ( $n = 28$ ).

The nucleotides involved in the B4-25 and CB-42 binding sites were determined by cloning and sequencing 20 members of each mutagenized and selected pool. As expected, the sequences were highly degenerate, containing an average of 9–10 substitutions in each 100-base sequence. This is smaller than the number of substitutions expected by chance (15) because of the conservation of about one quarter of the positions. The aligned aptamer sequences showed significant sequence conservation in specific regions (Fig. 3). For the number of sequences examined, only 3 or 4 bases would be expected to remain unmutagenized by chance. In fact, 28 positions were completely conserved in B4-25 and 22 in CB-42. The significance of the conserved sequences is supported by the fact that in each case an independent and otherwise unrelated clone from the selected pool (B4-49 and CB-45, respectively) shows sequence homology with the conserved region. A computer-generated secondary structure for the B4-25 sequence, along with the location of the conserved residues, is shown in Fig. 4.

To confirm that the selected sequences have an intact ligand binding site, RNAs were made from three variants from each pool and tested for binding (Table 3). Each bound with roughly the same affinity as the wild-type sequence. Thus, the conserved regions contain sequences that are likely to be involved in ligand binding.

### Discussion

We have shown that pools of roughly  $10^{13}$  different RNA molecules, of average length 100 nt, contain  $10^2$ – $10^5$  different

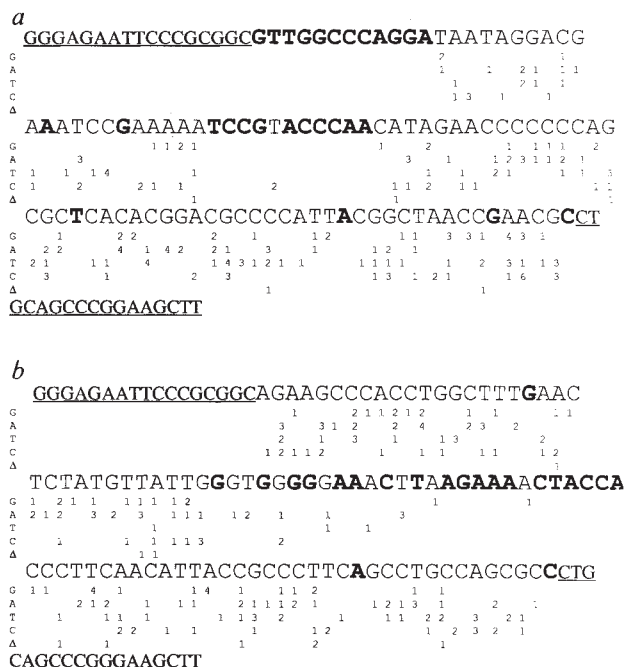


FIG. 3 Ligand binding sites within the sequences of the B4-25 (a) and CB-42 (b) aptamers. Invariant positions in the sequences of 20 clones derived from the mutagenesis and selection experiment described in the text are in bold; primer binding sequences are underlined. The number and type of mutations seen are indicated below each residue. As only 3 or 4 invariant positions would be expected in the absence of functional constraints, most of the invariant residues must contribute to ligand binding.

TABLE 3 Individual aptamers bind their cognate ligands

| Name of cloned aptamer              | Retention on cognate column (% binding) | Retention on Y column (%) | Mutational variant (cloned from mutagenized monoclonal) | Retention (%) |
|-------------------------------------|-----------------------------------------|---------------------------|---------------------------------------------------------|---------------|
| <b>Cibacron blue binding RNAs</b>   |                                         |                           |                                                         |               |
| CB-29                               | 49                                      | <1                        |                                                         |               |
| CB-30                               | 9                                       | 3                         |                                                         |               |
| CB-31                               | 65                                      | 80                        |                                                         |               |
| CB-37                               | 9                                       | <1                        |                                                         |               |
| CB-38                               | 25                                      | <1                        |                                                         |               |
| CB-41                               | 8                                       | —                         |                                                         |               |
| CB-42                               | 71                                      | 1                         | CB-42 wt                                                | 82            |
| CB-44                               | 0                                       | —                         | CB-42 No. 1                                             | 76            |
| CB-45                               | 64                                      | 2                         | CB-42 No. 9                                             | 69            |
| CB-46                               | 16                                      | <1                        | CB-42 No. 10                                            | 73            |
| Average of all CB sequences         | 32                                      |                           |                                                         |               |
| CB pool                             | 52                                      |                           |                                                         |               |
| <b>Reactive blue 4 binding RNAs</b> |                                         |                           |                                                         |               |
| B4-19                               | 40                                      |                           |                                                         |               |
| B4-21                               | 3                                       |                           |                                                         |               |
| B4-22                               | 44                                      |                           |                                                         |               |
| B4-25                               | 50                                      |                           | B4-25 wt                                                | 65            |
| B4-26                               | 13                                      |                           | B4-25 No. 1                                             | 64            |
| B4-27                               | 29                                      |                           | B4-25 No. 4                                             | 55            |
| B4-44                               | 25                                      |                           | B4-25 No. 5                                             | 69            |
| B4-45                               | 20                                      |                           |                                                         |               |
| B4-49                               | 42                                      |                           |                                                         |               |
| B4-51                               | 32                                      |                           |                                                         |               |
| B4-52                               | 53                                      |                           |                                                         |               |
| Average of all B4 sequences         | 32                                      |                           |                                                         |               |
| B4 Pool                             | 32                                      |                           |                                                         |               |

PCR DNAs from the sixth selection cycle of CB or B4 (or from the second selection cycle of mutagenized CB-42 or B4-25) were cloned into a derivative of pMLC28 (a gift of Dr Brian Seed) and sequenced. 'Monoclonal' DNA oligomers were produced by PCR amplification of individual plasmids and these oligomers were transcribed as described in Fig. 1 to produce single aptamers. After isolation, 2–5  $\mu$ g of RNA (in 150  $\mu$ l column buffer) was applied to small dye-ligand columns (1.25 ml bed volume) and these columns were developed as before (with appropriately smaller volumes), except that two column-volumes of 2 mM EDTA were used for elution. Fractions were counted and retention was determined by dividing the number of counts in the EDTA wash by the total number of counts applied. 'Retention on Y column' refers to cross-binding experiments between CB aptamers and Y-agarose; —, value not determined.

sequences capable of specific binding to a variety of small ligands. Because the total possible number of sequences 100-bases in length is  $10^{60}$ , even a pool of  $10^{13}$  molecules contains only a very small fraction of all possible sequences ( $10^{-47}$ ), and therefore the total number of different sequences that can bind to a given ligand must be in the range of  $10^{49}$ – $10^{52}$ . The question

of how many of these sequences represent truly independent solutions (that is, structurally different binding modes) is difficult to answer at this stage, but a rough estimate can be made. It is clear from our sequence data on RNAs from the CB and B4 pools that most of the binding RNAs bear no recognizable sequence similarity to each other. But we found one case from each pool of short regions of sequence similarity between two isolates, and binding site mapping by mutagenesis has implicated these sequences in the actual binding site. As the isolates that bear these conserved regions are otherwise entirely different, they must have arisen independently. This frequency of isolation of similar binding sites is consistent with the size of the two binding sites we mapped by mutagenesis: for example, CB-42 contains 22 conserved nucleotides. As there are only  $10^{13}$  different combinations of 22 nucleotides ( $4^{22}$ ), all such combinations should be present in a pool of  $10^{13}$  100-mers. The fact that the pool size is reduced to  $10^2$ – $10^5$  after selection, combined with the observed frequency of similar binding sites, implies that there may be no more than about 100–1,000 really different sequence solutions to the construction of specific binding sites for these ligands. It is important to note that there is no competition for binding under our selection system (ligand in excess), and that therefore our pools contain a variety of molecules with  $K_d$ s of less than about 500  $\mu$ M. Presumably the more tightly binding species are present at lower abundances.

Our results suggest that it may be possible to isolate novel ribozymes from pools of random-sequence RNAs. Some fraction of RNA molecules selected for binding to transition state analogue affinity columns would probably catalyze the corresponding reactions, by analogy with the isolation of catalytic antibodies selected for binding to transition state analogues

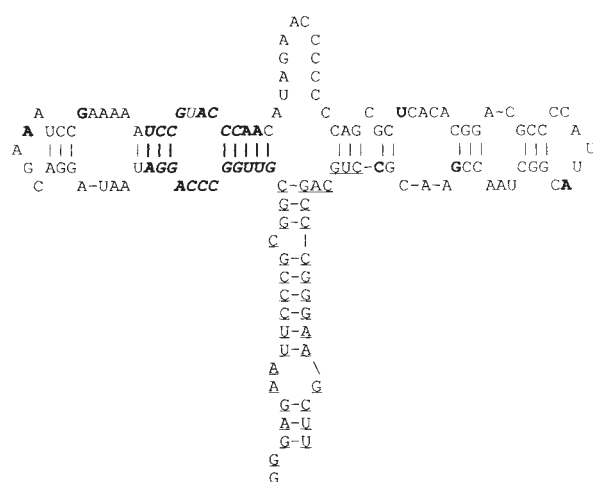


FIG. 4 Secondary structure diagram of the B4-25 aptamer generated by the FOLD program in the GCG sequence analysis package<sup>9</sup>. Primer binding sequences are underlined; the invariant residues that comprise the binding site are in bold, and those nucleotides in common with the B4-49 clone are in bold italics.



coupled to immunogenic substrates. Although it is difficult to extrapolate from substrate binding to catalytic function, the multiplicity of solutions seen for a given chemical problem

(ligand binding) suggests that complex catalysts such as the hypothesized primordial RNA replicase might also be accessible in even limited searches of sequence space. □

Received 29 May; accepted 20 July 1990.

1. Cech, T. R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4360–4363 (1986).
2. Orgel, L. E. *J. theor. Biol.* **123**, 127–149 (1986).
3. Benner, S. A., Ellington, A. D. & Tauer, A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7054–7058 (1989).
4. Thompson, R. T., Cass, K. H. & Stellwagen, E. *Proc. natn. Acad. Sci. U.S.A.* **72**, 669–672 (1975).
5. Arnold, F. H., Schofield, S. A. & Blanch, H. W. *J. Chromat.* **355**, 1–12 (1986).

6. Couture, S. *et al. J. molec. Biol.* (in the press).
7. Aruffo, A. & Seed, B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8573–8577 (1987).
8. Scharf, S. J., Horn, G. T. & Erlich, H. A. *Science* **233**, 1076–1078 (1986).
9. Devereux, J., Haeblerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).

ACKNOWLEDGEMENTS. This work was supported by Hoechst AG.

## LETTERS TO NATURE

### The Vela glitch of Christmas 1988

P. M. McCulloch, P. A. Hamilton, D. McConnell & E. A. King

Physics Department, University of Tasmania, GPO Box 252C, Hobart, Tasmania, Australia

IN the past ten years, the Vela pulsar PSR 0833–45 has undergone several large, discontinuous changes—glitches—in its pulsation period. On 24 December 1988, we were making continuous radio measurements of the Vela pulsar with a 2-min time resolution when a glitch occurred. Here we report our observations, which on the day of the glitch extend from 12 h before the event to 6 h after and represent the first time a glitch has been caught as it happened. The period decrease occurred without warning and in less than 2 min; an exponential recovery similar to previously observed post-glitch behaviour began immediately. Observations were made at 635 MHz and 950 MHz, and the lower frequency signal showed an additional delay starting at the time of the glitch and continuing for about 35 days. This behaviour is consistent with a small increase in dispersion measure or a change in the pulsar's magnetic field configuration. After the glitch, the arrival times of pulses at both frequencies differed from the smoothly predicted times according to a damped sinusoidal oscillation with a period of about 25 days.

In a typical Vela glitch, an abrupt decrease in period of about a part in a million is followed by a much slower recovery. Period measurements<sup>1,2</sup> taken within hours of each of the three most recent events show that an initial recovery, characterized by a slowing down on a timescale of a few days is followed by further recovery on a timescale of ~100 days. These observations are well described by a model in which the pre-jump pulse frequency  $\nu$  is given by a power series in time with terms in  $\nu$ ,  $\dot{\nu}$  and  $\ddot{\nu}$ , and the post-jump recovery consists of a constant-frequency offset plus two exponential recoveries.

The most recent period jump of the Vela pulsar occurred at ~19:00 UT on 24 December 1988 (Christmas Day at Hobart)<sup>3,4</sup>. Our observations on that day, part of a long-standing observational programme, were made simultaneously at 635 MHz and 950 MHz using a 14-m-diameter parabolic antenna with which the source is visible for 18 h a day. The receiver at each frequency is a full polarimeter, with dual-channel field-effect-transistor amplifiers receiving orthogonal linear polarizations which are combined to give the four Stokes parameters of the signal. The temperature corresponding to the system noise at each frequency is 60 K. The local oscillators are locked to a rubidium-vapour frequency standard which also drives the station clock. Receiver bandwidths of 250 kHz at 635 MHz and 800 kHz at 950 MHz limit the pulse broadening due to interstellar dispersion to <1% of the pulse period.

Integrated pulse profiles of 1,344 pulses from each polarimeter channel are recorded at intervals of ~2 min. The profiles are subsequently combined to give profiles of total intensity and

polarization information. The signal-to-noise ratio in each total intensity profile is typically 30:1, allowing a mean pulse arrival time to be determined to an accuracy of ~80  $\mu$ s at 635 MHz and ~50  $\mu$ s at 950 MHz. During every second integration a weak noise signal with known polarization properties is injected into the system to allow gain and phase calibration.

The data presented here were collected between 31 October 1988 and 27 March 1989. This interval was chosen because microglitches occurred immediately before and after these dates, complicating the analysis. The arrival time data have been reduced to arrival times at the Solar System barycentre using standard techniques. For these calculations the optical position of the Vela pulsar given by Manchester *et al.*<sup>5</sup>, precessed to the standard epoch of J2000.0 was used with data on the barycentre from the Jet Propulsion Laboratories ephemeris DE200.

For the interval from 31 October to 24 December, the pulse phase at time  $t$  for both frequencies is well described by

$$\phi(t) = \phi_0 + \nu(t - T_0) + \frac{1}{2}\dot{\nu}(t - T_0)^2 + \frac{1}{6}\ddot{\nu}(t - T_0)^3 \quad (1)$$

The parameters of this fit are given in Table 1a ( $\ddot{\nu}$  has a negative sign because the pulsar was recovering from a microglitch at JD 2447460). The mean value of the dispersion measure to the Vela pulsar derived from our data is 68.31443  $\text{cm}^{-3}$  pc. This is less than the value obtained some years ago, in accordance with a reported trend that the dispersion measure of this pulsar is decreasing with time<sup>6</sup>.

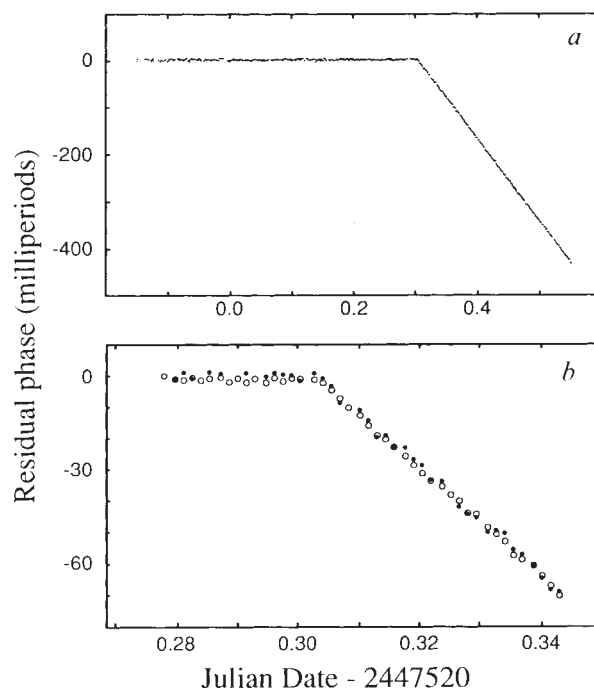


FIG. 1 The residuals from the fit using pre-jump parameters for the day of the glitch; each point represents a 2-min integration. *a*, 635-MHz data for the whole day. *b*, One hour's data including the glitch; ○, 950 MHz; ●, 635 MHz.

TABLE 1 Pre- and post-jump parameters for the Vela pulsar

|                                                                 |                                                  |
|-----------------------------------------------------------------|--------------------------------------------------|
| (a) Pre-jump parameters, $T_0 = \text{JD}^* 2447420.5519$       |                                                  |
| $\nu$                                                           | $11.19997850771 (2) \text{ Hz}$                  |
| $\dot{\nu}$                                                     | $-1.55792 (2) \times 10^{-11} \text{ Hz s}^{-1}$ |
| $\ddot{\nu}$                                                    | $-3.4 (4) \times 10^{-22} \text{ Hz s}^{-2}$     |
| (b) Parameters of the jump, $T_g = \text{JD} 2447520.30360 (8)$ |                                                  |
| $\Delta\nu$                                                     | $20.218 (9) \times 10^{-6} \text{ Hz}$           |
| $\Delta\dot{\nu}$                                               | $-1.2 (1) \times 10^{-12} \text{ Hz s}^{-1}$     |
| $\Delta\nu/\nu$                                                 | $1.805 \times 10^{-6}$                           |
| $\Delta\dot{\nu}/\dot{\nu}$                                     | $7.7 \times 10^{-2}$                             |
| (c) Parameters from fitting equation (3)                        |                                                  |
| $\Delta\nu_c$                                                   | $1.6662 (8) \times 10^{-5} \text{ Hz}$           |
| $\Delta\nu_1$                                                   | $1.086 (2) \times 10^{-7} \text{ Hz}$            |
| $\tau_1$                                                        | $4.64 (2) \text{ d}$                             |
| $\Delta\nu_2$                                                   | $3.396 (8) \times 10^{-6} \text{ Hz}$            |
| $\tau_2$                                                        | $351 (1) \text{ d}$                              |
| r.m.s. residual of fit                                          | $79 \mu\text{s}$                                 |

Numbers in parentheses are the  $2\sigma$  uncertainties in the last digit.

\* Julian date.

Figure 1a shows the residuals from the pre-jump fit for data taken on 24 December 1988. Shortly after 19:00 UT, the residuals diverge from the fit, indicating a sudden decrease in the pulse period. Figure 1b shows an hour of data starting at  $\sim 19:45$  UT on an expanded scale so that the individual points, each the result of a 2-min integration, are easily seen. The jump occurred on a timescale much shorter than 2 min. There is no warning of the impending glitch, and the observations are consistent with an instantaneous period change.

Data from the interval 23 to 26 December were fitted to a post-jump recovery function based on a two-component neutron-star model<sup>7</sup>

$$\nu(t) = \nu_p(t) + \Delta\nu_c + \Delta\nu e^{-(t-T_g)/\tau} \quad (2)$$

where  $T_g$  is the jump epoch,  $\nu_p(t)$  is the extrapolated pre-jump frequency,  $\Delta\nu_c$  is the permanent part of the frequency jump,

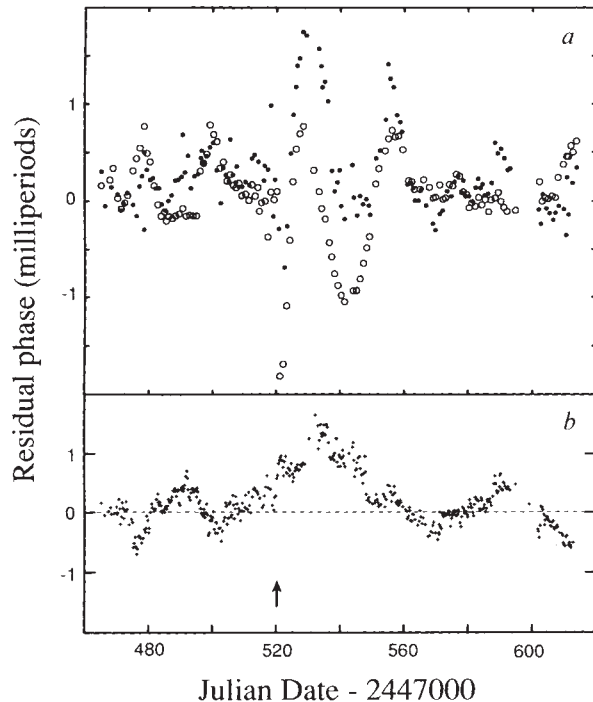


FIG. 2 *a*, The residuals from the fit using equation (3);  $\circ$ , 950 MHz;  $\bullet$ , 635 MHz. Each point is the average of one day's data. *b*, The phase difference obtained by subtracting the 950-MHz arrival times from the 635-MHz arrival times (positive value indicates that 635-MHz data arrived later). The arrow shows the epoch of the glitch.

and  $\Delta\nu$  is the component that decays with time constant  $\tau$ . Table 1b gives parameters derived from this fit.

The longer-term recovery from the jump was examined by analysing data to 27 March 1989. As for previous glitches the recovery was modelled by a function which included both long-term and short-term exponentials

$$\nu(t) = \nu_p(t) + \Delta\nu_c + \Delta\nu_1 e^{-t/\tau_1} + \Delta\nu_2 e^{-t/\tau_2} \quad (3)$$

The best-fit parameters for this model are given in Table 1c, assuming that the glitch occurred at JD 2447520.30360. The permanent frequency offset ( $\Delta\nu_c$ ) accounts for  $\sim 82\%$  of the frequency jump. The long-term exponential recovery (time constant 351 days) provides a further 17% and the short-term exponential recovery (time constant 4.6 days) accounts for the remainder.

Figure 2a shows the residuals from the complete jump function (equation (3)). Starting at about the jump epoch, there is an offset between the data at the two observing frequencies, with the 635-MHz data points consistently about one milliperiod later than the corresponding 950-MHz points. This offset disappears after  $\sim 40$  days. Figure 2a also shows a damped oscillation in the residuals with a maximum amplitude of  $\sim 2.5$  milliperiods and a period of  $\sim 25$  d occurring shortly after the jump. Neither of these effects have been reported before now.

Figure 2b shows the difference in arrival time between the two frequencies. The most significant feature is the prolonged positive excursion starting close to the epoch of the jump. All the curves show a series of quasi-periodic oscillations before the jump with a period of a few tens of days, similar to the damped sinusoid following the jump. We have examined the data for changes in mean flux, polarized flux or position angle of the linearly polarized component but found no effects.

The oscillatory behaviour preceding and following the jump is also apparent in the pulsar frequency derivative (Fig. 3). The recovery in  $\dot{\nu}$  towards its pre-jump value starts immediately following the jump and is approximately exponential; it does not exhibit the characteristic Fermi function shape predicted by Alpar *et al.*<sup>8</sup>.

One of the most interesting features is the time offset, apparent in Fig. 2, between the data at the two frequencies, which began

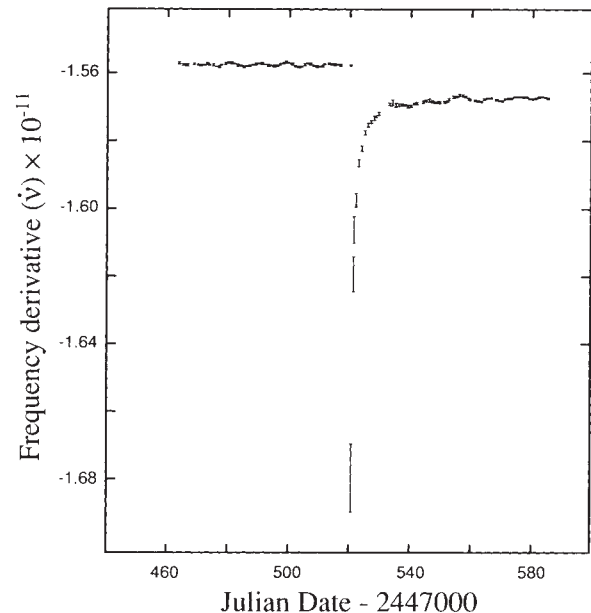


FIG. 3 The variation of frequency derivative  $\dot{\nu}$  with time. The error bars are  $\pm 2\sigma$ . Values were obtained by fitting for  $\nu$  and  $\dot{\nu}$  to sequences of 8 days of data, and advancing by one day between each fit. The values of  $\dot{\nu}$  immediately before and after the jump were calculated using the parameters given in Table 1.



near enough to the jump epoch to suggest a causal relationship. The data are too noisy to allow the exact time of this phenomenon to be determined, but suggest that it preceded the jump.

The impact of a comet on the neutron star could increase the dispersion measure, leading to the time offset, and could trigger a glitch. We would expect the comet to lose a significant fraction of its mass as it approached the pulsar; much of the mass lost would be ionized and would increase the dispersion measure. The required increase in column density is  $\sim 0.016 \text{ cm}^{-3} \text{ pc}$ , which represents a total mass of  $\sim 10^9 \text{ kg}$  if the material is spread in a spherical layer at the light cylinder radius. This is much less than the estimates of cometary masses by Harwit and Salpeter<sup>9</sup> who suggested that  $\gamma$ -ray bursts are the result of cometary impacts on neutron stars. The impact of the comet would cause a local rise in temperature which could trigger the period jump.

An alternative explanation is that the time offset reflects a change in the alignment of the pulsar magnetic field at the epoch of the jump. Such a change may be confined to regions of the magnetic field close to the surface of the pulsar and involve high-order multipole components. Similar non-dispersive phase

shifts have been observed in the high-frequency profiles from a number of pulsars, and interpreted in terms of magnetic field restructuring<sup>10</sup>.

Data at a third frequency would help to distinguish between these possibilities, and we have added a 1.4-GHz receiver to our 14-m telescope system.

The damped sinusoidal oscillation evident in the post-jump timing residuals is another new feature. A sinusoid in pulsar timing residuals is not uncommon, arising easily as an artefact of the model-fitting procedure. In such cases, however, the parameters of the sinusoid change as the model is varied. Here the damped sinusoid persists as the model is changed, and it also appears in the model-independent plot of  $\dot{\nu}$  (Fig. 3). This oscillation, with a period of  $\sim 25$  days, may be the result of a lightly damped Tkachenko oscillation, an axial vibration mode of the lattice of vortex lines in the rotating neutron superfluid<sup>11,12</sup>. The width of the equivalent cylindrical region would need to be  $\sim 1 \text{ km}$ . There is some evidence for a similar low-quality-factor oscillation in the timing residuals before the period jump (Fig. 2). If these are the result of a superfluid oscillation, the parameters should provide further constraints on models of the neutron superfluid in the neutron star.  $\square$

Received 23 April; accepted 14 June 1990.

1. McCulloch, P. M., Hamilton, P. A., Royle, G. W. R. & Manchester, R. N. *Nature* **302**, 319–321 (1983).
2. McCulloch, P. M., Klekociuk, A. R., Hamilton, P. A. & Royle, G. W. R. *Aust. J. Phys.* **40**, 725–730 (1987).
3. Flanagan, C. *IAU Circ.* 4695 (1989).
4. Hamilton, P. A., King, E. A., McConnell, D., McCulloch, P. M. *IAU Circ.* 4708 (1989).
5. Manchester, R. N. *et al. Mon. Not. R. astr. Soc.* **184**, 159–170 (1978).
6. Hamilton, P. A., Hall, P. J. & Costa, M. E. *Mon. Not. R. astr. Soc.* **214**, 57–6p (1985).

7. Baym, G., Pethick, C., Pines, D. & Ruderman, M. *Nature* **224**, 872–874 (1969).
8. Alpar, M. A., Anderson, P. W., Pines, D. & Shaham, J. *Astrophys. J.* **278**, 791–805 (1984).
9. Harwit, M. & Salpeter, E. E. *Astrophys. J.* **186**, L37–L39 (1973).
10. Kuzmin, A. D., Malofeev, V. M., Izvekova, V. A., Sieber, W. & Wielebinski, R. *Astr. Astrophys.* **161**, 183–194 (1986).
11. Ruderman, M. *Nature* **225**, 619–620 (1970).
12. Sonin, E. B. *Rev. mod. Phys.* **59**, 87–155 (1987).

## Fingering instability of thin spreading films driven by temperature gradients

A. M. Cazabat, F. Heslot, S. M. Troian\* & P. Carles

Physique de la Matière Condensée, Collège de France,  
11 place Marcelin Berthelot, 75231 Paris Cedex 05, France

THE dynamics of spreading of thin liquid films are important to many technological and biological processes, including tertiary oil recovery, coating processes, the formation and protection of microchips and biological cell interactions. The spontaneous spreading of thin liquid films under capillary forces alone is typically a slow process. An applied force—gravitational or centrifugal, or a surface shear stress—can be used to drive the spreading more quickly<sup>1,2</sup>. Recent experimental<sup>3,4</sup> and theoretical<sup>5</sup> studies have revealed that in all of these cases of forced spreading, the liquid front undergoes a fingering instability. Here we report another example of such unstable driven flow, this time caused by the Marangoni effect<sup>6</sup>, in which a temperature gradient induces a gradient of surface tension which drives the spreading process.

Thin films of liquids can be made to rise above the equilibrium meniscus position under the force of an upwardly directed gradient in the surface tension<sup>2</sup>. Such Marangoni films are formed by applying a thermal gradient along a vertical surface, which gives rise to a surface tension gradient along the direction of flow,  $d\gamma/dx = (d\gamma/dT)(dT/dx)$ , where  $\gamma$  is the local surface tension,  $T$  the local temperature and  $x$  the vertical distance along the plate. The temperature variation of the surface tension,  $(1/\gamma)(d\gamma/dT)$ , is fairly constant for many fluids far from the critical point, and therefore a constant temperature gradient creates a constant Marangoni surface stress  $\tau = d\gamma/dx$ . In the

lubrication approximation, the Navier–Stokes equations describing such unidirectional flow reduce to

$$\eta \frac{\partial^2 v}{\partial z^2} = \frac{\partial p}{\partial x} + \rho g \quad (1)$$

where  $v(x, z)$  is the fluid velocity upwards along the plate,  $z$  is the distance from the solid,  $\eta$  is the viscosity,  $p$  is the pressure in the fluid, and  $\rho g$  the gravitational force per unit volume. For very thin films,  $\rho g \ll \partial p/\partial x$ , and for small surface curvature  $\partial p/\partial x = -\gamma(d^3h/dx^3)$ , where  $h(x)$  defines the film surface. (To first order, one can ignore the temperature dependence of the viscosity, which induces higher-order effects than we investigate here, as well as the variation of surface tension in the pressure-gradient term.) Integrating equation (1) subject to the boundary conditions  $v(x, z=0) = 0$  and  $\eta(\partial v/\partial z)(z=h(x)) = \tau$  gives the height-averaged velocity

$$V(x) = \frac{h}{2\eta} \tau + \frac{\gamma}{3\eta} h^2 \frac{d^3h}{dx^3} \quad (2)$$

The first term describes the climbing of the film under Marangoni forces (for  $\tau$  positive), and the second term describes flow induced by surface curvature (Laplace pressure). Away from the sharply curved ends of the meniscus region near the bottom of the plate and the advancing front, the surface curvature is negligible and the fluid velocity is dominated by the first term. Under quasistatic conditions in which the incoming flux,  $Q = hV$ , is constant,  $h(x)$  will be nearly flat and the fluid film should spread linearly in time with a constant velocity  $V_{th}$  given by

$$V_{th} = \frac{h_0 \tau}{2\eta} \quad (3)$$

where  $h_0$  is the thickness of the flat film. This thickness is determined by both the detailed fluid geometry in the meniscus region and by the magnitude of the thermal stress pulling the film up along the wafer.  $h_0$  can be found by asymptotically matching the height profile from the Marangoni-driven region to the height profile determined from the meniscus region in which

\* Permanent address: Exxon Research and Engineering Corp., Clinton Twp., Route 22E, Annandale, New Jersey 08801, USA.

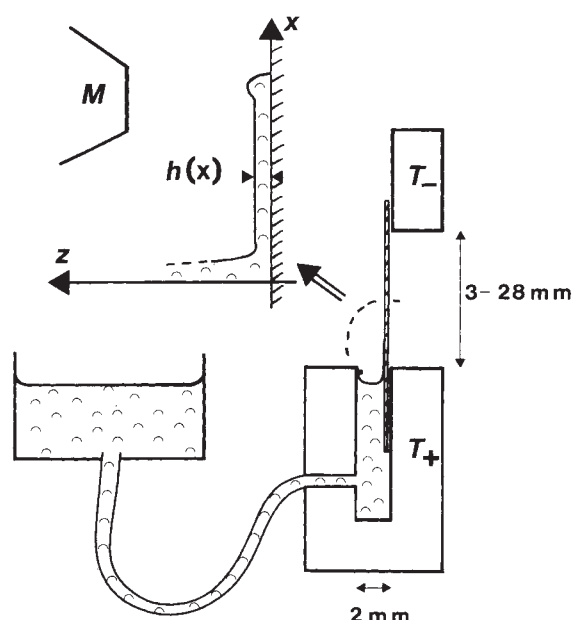


FIG. 1 A schematic of the capillary rise geometry. Upper left-hand corner shows a magnified view of the meniscus region.

only gravitational and capillary forces matter. This requires a knowledge of the specific boundary conditions near the injection site in the meniscus region. We need not solve explicitly for  $h_0$  in the analysis which follows.) The capillary contribution to the velocity becomes important near the advancing front where the fluid surface curves sharply to meet the substrate. Balancing the two terms in equation (2) gives that the length  $L$ , near the moving front, over which the capillary force is comparable to the viscous and Marangoni forces is

$$L = h_0(3Ca)^{-1/3} \quad (4)$$

where the capillary number is  $Ca = \eta V_{th}/\gamma$ . As first suggested by Huppert<sup>3</sup> and calculated explicitly in ref. 5, the length of this capillary region is what determines the wavelength  $\Lambda$  of the fingering instability, a result we confirm below.

TABLE 1 Dependence of spreading on viscosity and temperature gradient

| $\eta$<br>(mPa s) | $\tau$<br>(N m <sup>-2</sup> ) | $h_0$<br>( $\mu$ m) | $V_{exp}$<br>( $\mu$ m s <sup>-1</sup> ) | $\Lambda$<br>( $\mu$ m) | $V_{th}$<br>( $\mu$ m s <sup>-1</sup> ) | $L$<br>( $\mu$ m) | $\Lambda/L$ |
|-------------------|--------------------------------|---------------------|------------------------------------------|-------------------------|-----------------------------------------|-------------------|-------------|
| 20                | 0.5                            | 0.86                | 8*                                       | 600                     | 10.7                                    | 27                | 22          |
| 20                | 0.27                           | 0.65                | 3*                                       | 610                     | 2.8                                     | 27                | 22          |
| 20                | 0.21                           | 0.54                | 3                                        | 480                     | 2.8                                     | 25                | 19          |
| 20                | 0.10                           | 0.27                | 1                                        | 370                     | 0.68                                    | 20                | 18          |
| 20                | 0.054                          | 0.17                | 0.3                                      | 340                     | 0.23                                    | 18                | 19          |
| 100               | 0.21                           | 0.65                | 0.8                                      | 580                     | 0.7                                     | 28                | 21          |
| 500               | 0.21                           | 0.33                | 0.1†                                     | 340                     | 0.07                                    | 18                | 19          |

\* Linear regime too short or † slope not sufficiently constant to give a precise reading.

We have studied the climbing rate of films of a light silicone oil polydimethylsiloxane (PDMS, Petrarch Systems) on oxide-covered silicon wafers in a capillary-rise geometry as shown in Fig. 1. The wafers are cleaned with hexane, then methanol, providing a surface completely wettable by PDMS. The 250- $\mu$ m-thick wafers are held vertical by aspiration against two mobile brass blocks kept at a constant temperature difference of 28 °C (top block colder). We estimate the surface tension gradient in the film by calculating the temperature gradient along the wafer (28 °C divided by the distance between the blocks).

Fluid enters the system through a plastic tube connecting the reservoir bottle and the pool of liquid in which the bottom edge of the wafer is dipping. The temperature gradient is first established with the wafer in position; then the level of fluid is slowly raised (by moving the reservoir bottle) to where the fluid meniscus on the wafer just reaches into the thermal gradient region.

The climbing film is observed through a reflection microscope with laser illumination ( $\lambda = 6,328 \text{ \AA}$ ). Equal-thickness fringes, spaced  $\lambda/2n$  apart (where  $n = 1.4$  is the optical index of the PDMS), are used to reconstruct the thickness profiles. Figure 2 *a-d* shows the typical development of a film spreading under a temperature gradient. The spurious vertical interference lines correspond to fringes from a half-silvered mirror inclined at 45 ° in the microscope and used for illumination. Some time after the liquid is put into contact with the vertical wafer (the time delay depending on the fluid viscosity and the magnitude of the temperature gradient), a film of relatively constant thickness begins climbing the wafer (Fig. 2*a*). A bump in the height profile,

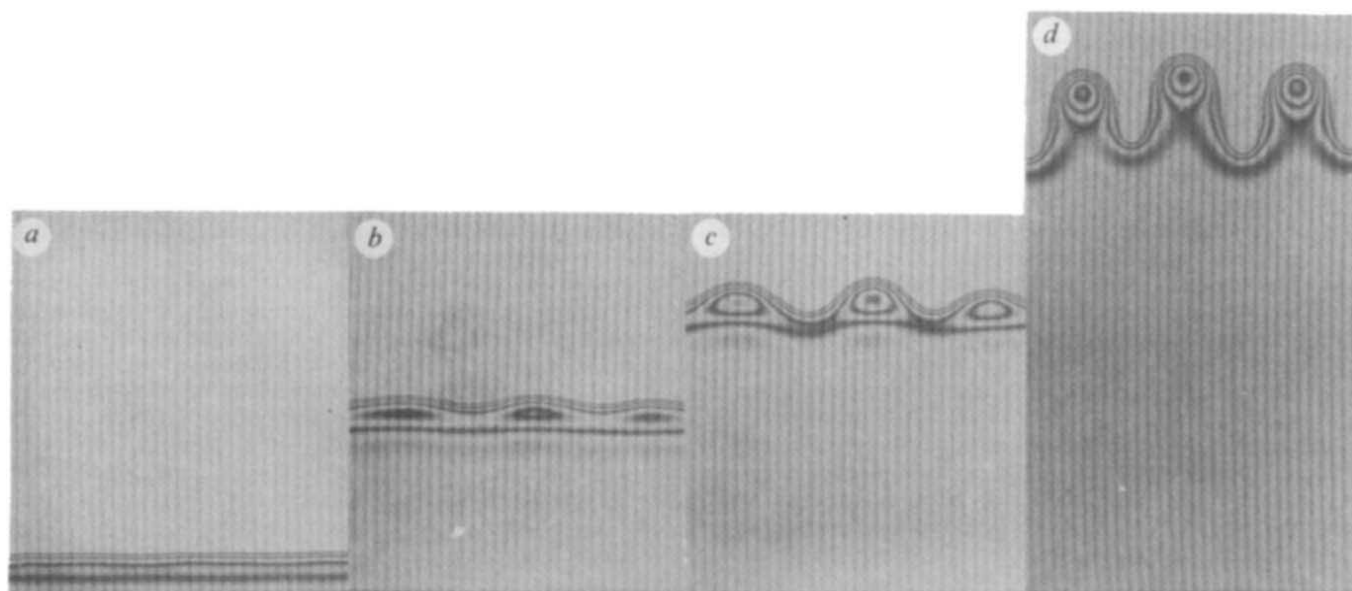


FIG. 2 Time development of the fingering instability at the vertically ascending front of a PDMS oil film of 20 mPa s<sup>-1</sup> and  $\tau = 0.18 \text{ Pa}$ . *a*, 1.5 min; *b*,

6.5 min; *c*, 10 min; *d*, 17 min. Measured from tip to tip, the wavelength  $\Lambda$  in *d* is 0.5 mm.



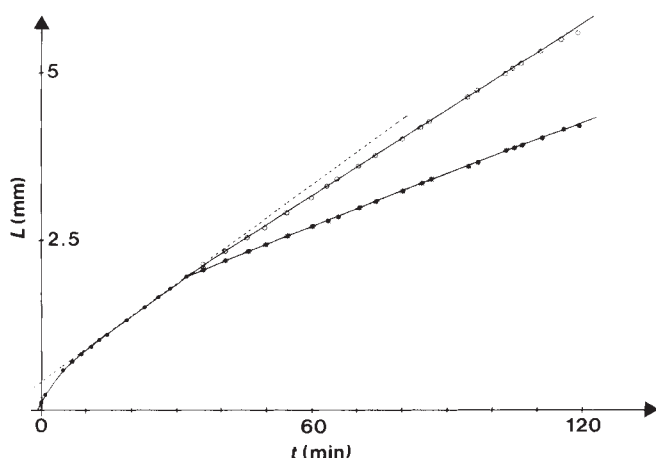


FIG. 3 Position of the fluid front as a function of time for PDMS of  $100 \text{ mPa s}^{-1}$  and  $\tau = 0.21 \text{ Pa}$ . The values  $V_{\text{exp}}$  in Table 1 are determined from the slope of the linear regime occurring before the onset of the instability.

indicated by the closely spaced fringes near the front of the film, develops near the climbing front (Fig. 2b), which soon becomes unstable (Fig. 2c) and results in a fingering instability with a rather well defined wavelength (Fig. 2d). Figure 3 shows the position of the fluid front as a function of time. The fork in the curve signals the beginning of the instability, the upper curve corresponding to the finger tips and the bottom one to the 'valley'. After some initial transients, but before the onset of the instability, the front moves linearly in time, as predicted for a film of constant height. Once the instability has developed, both the tips and valleys also grow linearly in time. Not surprisingly, the linear development for the finger growth is different both from that for gravitational<sup>3</sup> and for rotational flow<sup>4</sup> because of the absence of constant fluid volume and different flow geometry (linear instead of radial) in our case.

For more detailed comparison between the predicted and measured values of the film velocity without the fingers and of the wavelength of the instability, we used silicone oils of varying viscosity and applied several different temperature gradients as shown in Table 1. The quantities  $h_0$  and  $V_{\text{exp}}$  are experimentally measured values.  $\Lambda$  is the measured wavelength of the instability, and  $V_{\text{th}}$  and  $L$  were calculated using equations (3) and (4). For the oils used, the surface tension,  $\gamma = 20 \text{ mN m}^{-1}$ , does not vary much with molecular weight, and  $|(1/\gamma)(d\gamma/dT)| = 0.0025 \text{ K}^{-1}$ . In evaluating the capillary number, we used the viscosities of the fluids at  $25^\circ\text{C}$ . The agreement between the measured and predicted values of the velocity is quite good.

The height profiles in the absence and presence of the instability as well as the growth rate of the instability as a function of wavenumber have been calculated in a linear stability analysis (E. Herbolzheimer and S. M. Troian, manuscript in preparation). A detailed comparison of the predicted and measured height profiles will be published elsewhere, but this analysis shows that the pronounced bump observed in the height profile causes a capillary-driven fingering instability more complex yet not unlike the Rayleigh instability (the breakup of a liquid column into droplets). In our case, one can think of the bump as a long roll of fluid (in the  $y$  direction) the axial symmetry of which is broken by liquid being pumped in from the back. The bump forms because the contact line cannot move as fast as the fluid being pumped in from behind by the Marangoni stress. When the bump or roll develops enough axial curvature, the timescale depending on the thermal gradient and the liquid viscosity, the roll becomes unstable and breaks up into fingers. If the mode observed is the one with the fastest-growing wavelength, the theoretical ratio of  $\Lambda/L = 25$  is in good agreement with the average measured value of 20 (Table 1). □

Received 17 April; accepted 16 July 1990.

1. Tanner, L. H. *La Recherche* **17**, 182–191 (1986).
2. Ludviksson, V. & Lightfoot, E. N. *Am. Inst. Chem. Engrs J.* **17**, 1166–1173 (1971).
3. Huppert, H. E. *Nature* **300**, 427–429 (1982).
4. Melo, F., Joanny, J. F. & Fauve, S. *Phys. Rev. Lett.* **63**, 1958–1961 (1989).
5. Troian, S. M., Herbolzheimer, E., Safran, S. A. & Joanny, J. F. *Europhys. Lett.* **10**, 25–30 (1989).
6. Levich, V. G. & Krylov, V. S. *A. Rev. Fluid Mech.* **1**, 293 (1969).

## Significant flux of atmospheric nitrous oxide from the northwest Indian Ocean

C. S. Law & N. J. P. Owens

Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, Devon PL1 3DH, UK

INTEREST in nitrous oxide ( $\text{N}_2\text{O}$ ) has increased considerably in the light of its deleterious effect on the ozone layer<sup>1</sup>, and contribution to the greenhouse effect<sup>2</sup>. There are many sources of atmospheric  $\text{N}_2\text{O}$ , both anthropogenic (for example, combustion) and natural, but the global budget is still inadequately defined. Despite the fact that most of the world's oceans are close to equilibrium with the atmosphere<sup>3</sup>, water bodies depleted in oxygen have been identified as areas of  $\text{N}_2\text{O}$  production<sup>4,5</sup>, and so the oceans represent a potential source of atmospheric  $\text{N}_2\text{O}$ . Here we report elevated  $\text{N}_2\text{O}$  concentrations in the northwest Indian Ocean, an area that exhibits upwelling and high oxygen consumption in the water column. We found that  $\text{N}_2\text{O}$  was supersaturated in both oxygen-saturated surface waters (up to 246%  $\text{N}_2\text{O}$  saturation) and oxygen-depleted sub-surface waters (1,264%  $\text{N}_2\text{O}$  saturation). The calculated flux to the atmosphere indicated that upwelling in the northwest Indian Ocean ( $15\text{--}25^\circ\text{N}$ ) represents one of the most significant marine sources of  $\text{N}_2\text{O}$ , contributing between 5 and 18% of the total marine flux from a surface area of only 0.43% of the world ocean. These data suggest that the oceanic flux of  $\text{N}_2\text{O}$  to the atmosphere shows strong spatial heterogeneity which should be considered in global budgets and ocean-atmosphere models.

A transect of the northwest Indian Ocean (Fig. 1) was followed from close to the Equator ( $00^\circ04'\text{N}$   $65^\circ05'\text{E}$ ), across the Arabian Sea into the Gulf of Oman ( $24^\circ49'\text{N}$   $57^\circ13'\text{E}$ ) during September–October 1986. Vertical profiles of dissolved  $\text{N}_2\text{O}$  (and other oceanographic variables) were obtained at 16 stations by sampling at 20–24 depths between the surface and 4,000 m. This region exhibits two important oceanographic features: (1) a monsoon-driven upwelling, located off the Arabian Peninsula; (2) sub-surface waters that are persistently severely depleted in oxygen<sup>6</sup>. The upwelling is dominant during June–August when the southwesterly winds average  $10\text{--}11 \text{ m s}^{-1}$ ; whereas the rest of the year is characterized by a mean wind speed of  $4 \text{ m s}^{-1}$  (Meteorological Office, Bracknell, UK) with reduced upwelling. The water column was generally supersaturated with  $\text{N}_2\text{O}$  with respect to an atmospheric  $\text{N}_2\text{O}$  concentration of 303.9 parts per  $10^9$  by volume (ref. 3, with subtraction of 0.2% to account for the annual increase in atmospheric  $\text{N}_2\text{O}$ ); the data are therefore considered in terms of  $\Delta\text{N}_2\text{O}$ , which represents the excess  $\text{N}_2\text{O}$  over that which would be in equilibrium with the atmosphere. Dissolved oxygen concentrations are presented in terms of the apparent oxygen utilization, AOU (the difference between the equilibrium oxygen concentration at ambient temperature and salinity and the observed oxygen concentration), which is a measure of the degree of oxygen depletion.

Figure 1 shows the  $\text{N}_2\text{O}$ -depth distribution from stations 1 to 11. All profiles exhibited a similar trend with an increase in  $\Delta\text{N}_2\text{O}$  between 200 and 1,000 m, coincident with the oxygen minimum.  $\Delta\text{N}_2\text{O}$  maxima of  $>45 \text{ nmol l}^{-1}$  were recorded in the upwelling region (stations 6–16), and also at station 3 where lower-salinity surface water, originating from the region of the

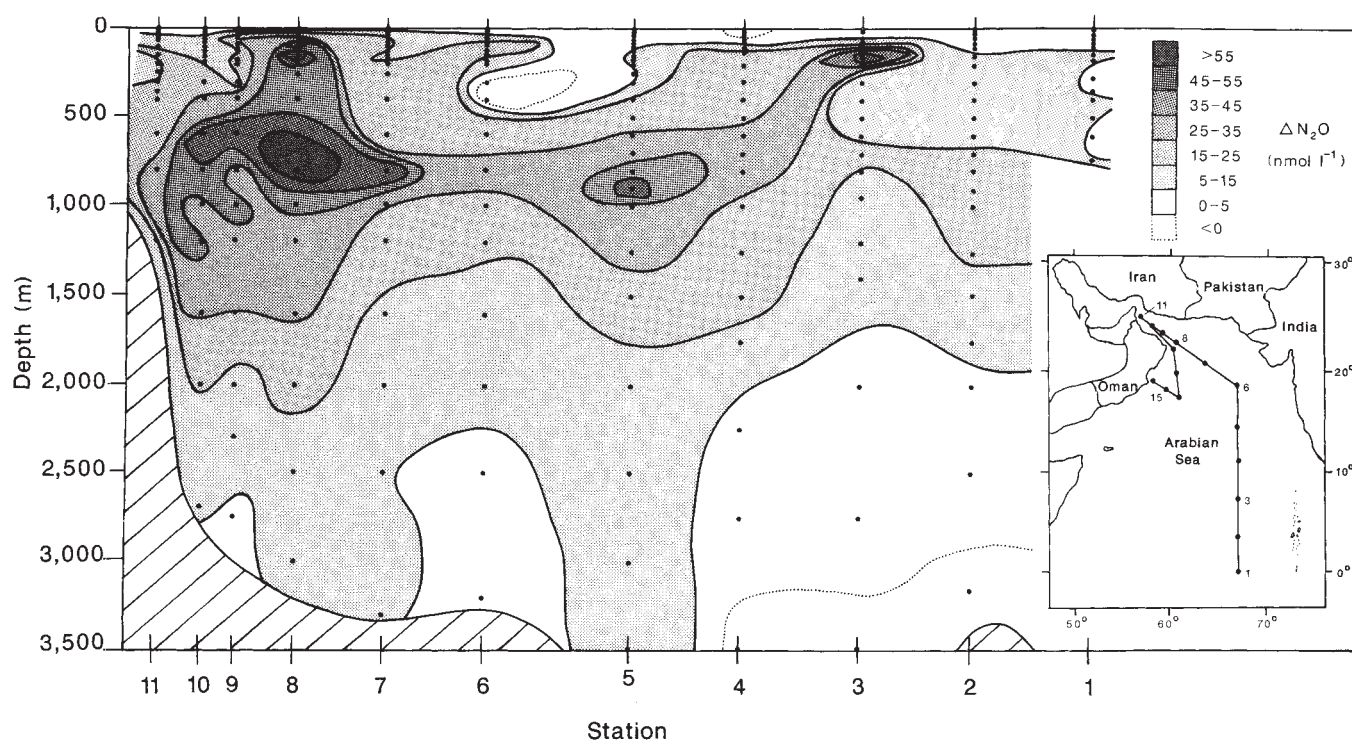


FIG. 1  $\Delta N_2O$  depth distribution along a transect from station 1 at the Equator to station 11 in the Gulf of Oman (see inset). Water samples (20–24 depths from the sea surface to the bottom) were obtained by Go-Flo oceanographic bottles and subsampled immediately under nitrogen. Dissolved  $N_2O$  was extracted immediately by a static vacuum technique<sup>21</sup>. In brief, water samples were drawn into an evacuated 500-ml gas sampling bulb through a narrow constriction, which facilitated the dissociation of the gas and liquid phase. After attaining atmospheric pressure, 1-ml aliquots of the headspace were removed and analysed by electron capture detector/gas chromatogra-

phy (Varian, 2 m  $\times$  3 mm outer diameter Porapak Q column, with a precolumn of drierite and carbosorb; column at ambient temperature, detector 320 °C; high-purity  $N_2$  carrier at flow rate 30 ml min<sup>-1</sup>). Calibration and efficiency of the extraction technique was determined by analysing sea water of known  $N_2O$ . The standard deviation of the gas chromatographic analysis was 1.4%, and the relative standard deviation of the entire analysis was 3.8%.  $N_2O$  (99.5% purity) for calibrations obtained from Bedfont Technical Instruments Ltd.

Equator, dipped below water of elevated salinity originating from the Gulf of Oman. The average  $\Delta N_2O$  supersaturation of all samples ( $n = 309$ ) was 302%, with a maximum of 104.1 nmol l<sup>-1</sup> (1,264%) at a depth of 800 m at station 15.

Two areas of  $N_2O$  undersaturation were also observed; one at a depth of 350 m at station 6, concomitant with a nitrite maximum, and the other associated with Antarctic bottom water near the Equator. The coincidence of the  $\Delta N_2O$  minimum with a  $NO_2^-$  maximum has been noted previously<sup>4,7,8</sup>, and may be attributed to competition between denitrifiers and other bacteria for nitrate in the presence of an enhanced organic-matter supply<sup>9</sup>, or other factors<sup>10</sup>.

Figure 2 shows the relationship between  $\Delta N_2O$  and AOU. A positive correlation between these variables has been reported from a number of ocean regions<sup>9,11</sup> and, because AOU is highly correlated with nitrate regeneration<sup>12</sup> (bacterial nitrification), it has been concluded that nitrification is responsible for  $N_2O$  production<sup>4,13</sup>. Statistical analysis of our  $\Delta N_2O$ -AOU data indicates that a single linear relationship does not fully explain the data. The inclusion of temperature as a variable does not improve the relationship, in contrast to observations in the east equatorial Pacific Ocean<sup>5</sup> and the west Pacific/east Indian Ocean<sup>3</sup>. Instead, two lines provide the best fit to the relationship (37% of the variability explained), with a change in gradient at an AOU value of 197  $\mu\text{mol l}^{-1}$ . This has not been previously reported and may indicate that a single process may account for  $N_2O$  production, in which the rate increases significantly under increasingly oxygen-depleted conditions. Alternatively, a second process, with a greater potential for  $N_2O$  production may occur at low oxygen concentrations. The first postulate supports the view that nitrification is largely responsible for  $N_2O$  production

because, although predominantly an aerobic process,  $N_2O$  production by nitrification can increase under suboxic conditions<sup>14</sup>. The second postulate supports the view that an anaerobic process (such as denitrification) is largely responsible for the  $N_2O$  production<sup>15</sup>. The  $N_2O$  produced at high oxygen concentrations

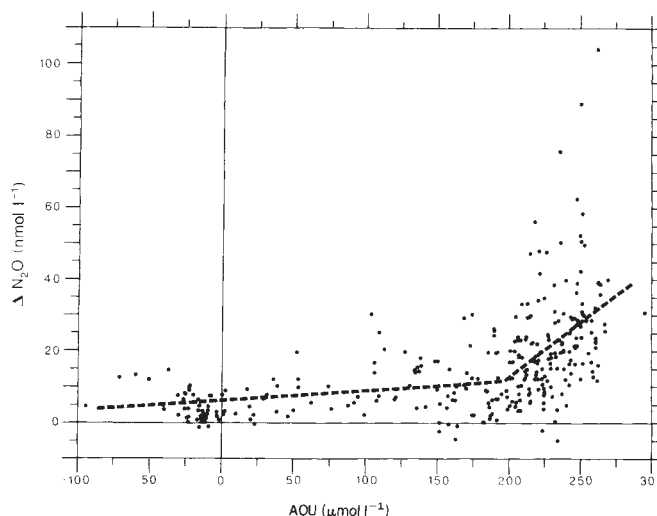


FIG. 2  $\Delta N_2O$ -apparent oxygen utilization (AOU) relationship for the northwest Indian Ocean. The data set is best described (37% variability explained) by two linear regressions with an inflexion at 197  $\mu\text{mol l}^{-1}$  AOU. The individual regressions are: for AOU < 197  $\mu\text{mol l}^{-1}$ ,  $y = 5.5 + 0.033x$ ,  $n = 142$ , significant at the 1% level; for AOU > 197  $\mu\text{mol l}^{-1}$ ,  $y = -49.4 + 0.31x$ ,  $n = 165$ , significant at the 1% level.



may be accounted for by both nitrification and denitrification in the microanaerobic centres of suspended particles<sup>16,17</sup>. These data cannot distinguish which process is most active in producing  $\text{N}_2\text{O}$  in the northwest Indian Ocean, but suggest that partitioning of the processes may occur as a function of the ambient oxygen concentration. Both nitrification and denitrification are responsible for  $\text{N}_2\text{O}$  production in the western north Pacific Ocean<sup>16</sup>.

Recent large-scale studies have shown that the ocean is predominantly in equilibrium with the atmosphere, with reported  $\text{N}_2\text{O}$  surface saturations of less than 102.5% (ref. 3) and 104% (ref. 18). The average surface saturation for the northwest Indian Ocean was  $167 \pm 46\%$  ( $\Delta\text{N}_2\text{O} = 3.47 \pm 2.48 \text{ nmol l}^{-1}$ ), indicating that this area is a significant source of atmospheric  $\text{N}_2\text{O}$ . To calculate the flux to the atmosphere, the study area was divided into two regions (0–15° N and 15–25° N), to account for the localization of the upwelling. Although there will be some variability associated with the seasonal intensity of the upwelling, the data in this study were obtained during the decline of the southwest monsoon, and so are representative for annual flux calculations. The atmospheric flux for the upwelling region (15–25° N), was first calculated from a  $\Delta\text{N}_2\text{O}$  value of  $4.49 \text{ nmol l}^{-1}$  ( $\pm 2.21$  [186.5  $\pm$  40.4%]) using the stagnant-film model, which has been used in other  $\text{N}_2\text{O}$  flux calculations<sup>9,11</sup>. Using a diffusion coefficient of  $2.35 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$  (from an average surface temperature of 27.7 °C) and a stagnant film thickness  $z$  of 59  $\mu\text{m}$  (calculated from an average annual wind speed of 5  $\text{m s}^{-1}$  (Meteorological Office, Bracknell, UK) and radon data from the Atlantic and Pacific Oceans<sup>9</sup>), we obtained an atmospheric flux of  $0.79 \pm 0.39 \text{ pg N}_2\text{O cm}^2 \text{ s}^{-1}$ . We note that the value of  $z$  used was significantly higher than used previously for  $\text{N}_2\text{O}$  flux calculations in the equatorial region<sup>9</sup>, and therefore our estimate is conservative.

To improve the precision of the atmospheric flux estimate, we used a second model. Different techniques have shown, from theoretical and field data, that the flux of a gas across the air/sea interface is a function of the windspeed. Assuming that the transfer velocity of  $\text{N}_2\text{O}$  will be similar to that of  $\text{CO}_2$ , owing to their identical molecular weights, we calculated a transfer-velocity range of 5–11.1  $\text{cm h}^{-1}$  for  $\text{N}_2\text{O}$ . This was obtained from a compilation of global data, including estimates of the transfer velocity of  $\text{CO}_2$  at 20 °C from radon and  $\text{SF}_6$  data incorporating the model equations in refs 19, 20. The range of transfer velocity used was significantly lower than that of 9.2–17.9  $\text{cm h}^{-1}$  calculated for upwelling in the east Indian Ocean<sup>3</sup>, because of our use of the average annual wind speed (5  $\text{m s}^{-1}$ ), and therefore our estimates are again conservative. Using a mean transfer velocity of 8.05  $\text{cm h}^{-1}$ , a flux of  $0.44 \pm 0.22 \text{ pg N}_2\text{O cm}^2 \text{ s}^{-1}$  was obtained, which is lower than that estimated from the stagnant-film model.

Using the results of these two methods for the range of  $\text{N}_2\text{O}$  flux, and a surface area of  $1.56 \times 10^6 \text{ km}^2$  for the upwelling region (15–25° N 50–72° E), we calculate an annual flux of  $\text{N}_2\text{O}$  to the atmosphere of 0.22–0.39 Tg  $\text{N}_2\text{O}$ . This does not include  $\text{N}_2\text{O}$  production in the upwelling off the Somali coast, and therefore underestimates the total atmospheric  $\text{N}_2\text{O}$  flux from the northwest Indian Ocean. If our annual flux is compared to a global marine flux of  $<4.4 \text{ Tg N}_2\text{O yr}^{-1}$  (ref. 3) then at least 5–9% may originate from the upwelling region. A recent survey carried out in 1987 over a large area suggested, however, that the annual marine flux may be  $<2.2 \text{ Tg N}_2\text{O yr}^{-1}$  (ref. 3), and the current study area would then account for at least 10–18% of the flux. Thus, our data indicate that a quantitatively significant proportion (5–18%) of the total marine source of  $\text{N}_2\text{O}$  is concentrated in a small proportion (0.43%) of the world ocean (a natural chimney?). This may have an important localized effect on atmospheric concentrations and implications for atmospheric chemistry. Further, the estimates for the total marine source of  $\text{N}_2\text{O}$  may be underestimates, which would have consequences for ocean-atmosphere climate models. □

Received 2 April; accepted 19 July 1990.

- Hahn, J. & Crutzen, P. J. *Phil. Trans. R. Soc.* **296**, 521–541 (1982).
- Yung, Y. L., Wang, W. C. & Lacis, A. A. *Geophys. Res. Lett.* **3**, 619–621 (1976).
- Butler, J. H., Elkins, J. W., Thompson, T. M. & Egan, K. B. *J. geophys. Res.* **94**, 14865–14877 (1989).
- Elkins, J. W., Wofsy, S. C., McElroy, M. B., Kolb, C. E. & Kaplan, W. E. *Nature* **275**, 602–605 (1978).
- Cline, J. D., Wisegarver, D. P. & Kelly-Hansen, K. *Deep Sea Res.* **34**, 857–873 (1987).
- Qasim, S. Z. *Deep Sea Res.* **29**, 1041–1068 (1982).
- Cohen, Y. & Gordon, L. I. *Deep Sea Res.* **25**, 509–524 (1978).
- Rasmussen, R. A. & Pierotti, D. *Pure appl. Geophys.* **116**, 405–413 (1978).
- Hahn, J. *Denitrification, Nitrification and Atmospheric Nitrous Oxide* (ed. Delwiche, C. C.) 191–240 (Wiley-Interscience, New York, 1981).
- Codispoti, L. A. *et al. Science* **233**, 1200–1202 (1986).
- Cohen, Y. & Gordon, L. I. *J. geophys. Res.* **84**, 347–353 (1979).
- Redfield, A. C., Ketchum, B. H. & Richards, F. A. *The Sea* (ed. Hill, M. N.) 26–77 (Wiley-Interscience, New York, 1963).
- Yoshida, N., Hattori, A., Saino, T., Matsuo, S. & Wada, E. *Nature* **307**, 442–444 (1984).
- Goreau, T. J. *et al. Appl. envir. Microbiol.* **40**, 526–532 (1980).
- Jørgensen, K. S., Jensen, H. B. & Sørensen, J. *Can. J. Microbiol.* **30**, 1073–1078 (1984).
- Yoshida, N. *et al. Nature* **342**, 895–897 (1989).
- Law, C. S. & Owens, N. J. P. *Neth. J. Sea Res.* **25**, 65–74 (1990).
- Weiss, R. F. *Eos* **110**, 1101–1102 (1978).
- Upstill-Goddard, R. C., Watson, A. J., Liss, P. S. & Liddicoat, M. I. *Tellus* (in the press).
- Liss, P. S. & Merlivat, L. *The Role of Air-Sea Exchange in Geochemical Cycling* (ed. Buat-Menard, P.) 113–127 (Riedel, Dordrecht, 1986).
- Law, C. S. thesis, Univ. Dundee (1989).

ACKNOWLEDGEMENTS. We thank J. L. Morris, M. R. Carr, P. H. Burkill and A. J. Watson, for advice and assistance, and acknowledge the help of colleagues, officers and crew of the RRS *Charles Darwin*. C.S.L. was the holder of an NERC studentship.

## Evidence for the Hudson River as the dominant source of sand on the US Atlantic Shelf

Dennis A. Darby

Department of Geological Sciences, Old Dominion University, Norfolk, Virginia 23529–0496, USA

PREVIOUS attempts to determine the origin of sand on the US Atlantic Shelf by standard heavy mineral analysis have concluded that local rivers and coastal erosion were probable sources<sup>1</sup>. Here a new approach to determine the history of surficial sand dispersal is developed using the elemental composition of detrital Fe–Ti oxide grains (primarily ilmenite) instead of heavy mineral abundances. Sands from the entire middle Atlantic Shelf (latitude 35° N to 41° N) have been interpreted as coming from specific drainage basins<sup>1,2</sup>. More than half of the sand samples collected as far south as North Carolina can be traced to the Hudson River and other New England sources. The dominance of the Hudson River is cause for rethinking sources of shelf sediment south of New Jersey. These findings suggest that the Hudson River may have rivalled the sediment discharge of much larger rivers like the Mississippi or Amazon Rivers at least during brief glacial melting events.

The use of geochemical data from the ilmenite–haematite–titanomagnetite opaque minerals avoids several problems commonly encountered in provenance studies<sup>3</sup>. For example, selective removal by winnowing is greatly reduced because of the relatively narrow density range of this group<sup>4</sup>. Also, these Fe–Ti oxide minerals, especially ilmenite ( $\text{FeTiO}_3$ ), are more ubiquitous than most heavy minerals, occurring as accessory minerals in a wide variety of igneous and metamorphic rocks<sup>5</sup>. Previous heavy-mineral studies have shown that only rare honey-yellow axinite grains found on the outer shelf of the middle Atlantic bight could be traced to specific sources in New Jersey<sup>6</sup>. Ilmenite, however, is one of the four most common heavy minerals in sands from the Atlantic Shelf and its largest source rivers<sup>7–9</sup>. Yet, significant differences in the major- and minor-element composition of these oxides result from the chemical conditions during crystallization of the source rock<sup>10,11</sup>.

Differences in the 10 measured elements distinguish the potential sand sources from one another (Table 1). Discriminant function analysis of the ilmenite chemistry correctly classifies all 63 samples from the Roanoke, James, Potomac, Susquehanna

and Hudson Rivers and the Long Island Shelf area (Fig. 1) into the source groups to which they belong at better than 0.95 probability (Fig. 2). The 11 Long Island Shelf samples, located northeast of the Hudson River palaeodrainage system across the shelf, would be representative of any sand moving south into the study area from the New England shelf, an area where the sand source is dominated by earlier glacial outwash from the adjacent land<sup>12</sup>. The samples from the centre of the Chesapeake Bay represent Susquehanna River sand. The ilmenite composition of these mid-bay samples is similar to samples from the Susquehanna River and different from Potomac River and other nearby sands.

The discriminant function is calculated using the element composition of the Fe-Ti oxide grains from these six potential source groups. This function assigns each of the 114 shelf samples to the most probable source group. If a shelf sample is classified as belonging to a source group at <0.95 probability, a multiple source is assumed consisting of the two most probable group memberships as designated by the discriminant function. Because individual grains were not analysed, the true proportion of grains belonging to multiple sources cannot be determined. Fortunately 72% of the shelf samples can be attributed to a single source with a probability >0.95. Even these 82 samples have some contribution from other sources but the bulk ilmenite composition was not significantly altered by this mixing.

Assuming that the other detrital grains have essentially the same sources as the Fe-Ti oxides, the dominant influence of the Hudson River on the Atlantic Shelf from New York to

Hatteras is clearly shown by this discriminant-function classification (Table 2). Of the total 114 shelf samples, the Hudson River is the source for 52% of the samples. Only 12 samples (10.5%) can be classified solely with sources south of the Hudson River. This disparity cannot be explained by a lack of ilmenite in these other rivers. The Roanoke and James Rivers, for example, have slightly higher abundances of this mineral than the Hudson River. The most reasonable explanation for the dominance of Hudson River Fe-Ti oxide grains on the shelf is a much larger volume of sand discharged to the shelf from the Hudson River relative to the rivers farther south.

The Hudson River sand is the dominant source in every shelf area except the area south of Cape Hatteras (area 4, Fig. 1). The probability that area 4 samples that classified with the Hudson River or Long Island Shelf sources also received some contribution from an unidentified source was fairly large (0.5) (Table 2).

The only major unsampled potential source is the Delaware River drainage basin. This potential source river could not drastically alter the classification results of Table 2. Only 22% of the samples in shelf area 2, the area most influenced by the Delaware River, had a significant contribution of grains from unsampled sources. The fact that seven of these ten samples were within 30 km of the mouth of the Delaware Bay suggests that the Delaware River might have contributed to these samples. This possibility is supported by the elevated TiO<sub>2</sub> content (45–62%) of these seven samples and the high TiO<sub>2</sub> ilmenite in the Delaware River drainage basin<sup>13</sup>. Thus most of the remaining

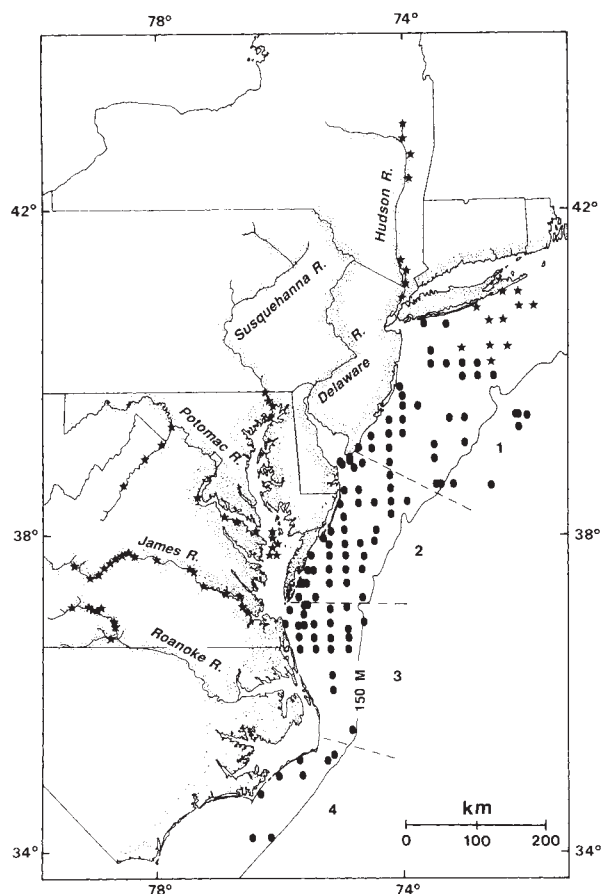


FIG. 1 Sample distribution on the US Atlantic Shelf (circles) and sample sites from potential sources (stars). The dashed boundaries between shelf areas 1–3 approximate depositional compartments between the principal estuaries and their offshore palaeodrainage patterns. Area 4 is south of the Hatteras promontory at which point the shelf is narrowest. River samples were taken from the unvegetated banks and mid-channel bars and occasionally from river channels using a grab sampler.

TABLE 1 Mean oxide content

(a) Potential Sources

|                                | Roanoke<br>River<br>(n=8) | James<br>River<br>(n=20) | Potomac<br>River<br>(n=8) | Susquehanna<br>River<br>(n=8) | Hudson<br>River<br>(n=8) |
|--------------------------------|---------------------------|--------------------------|---------------------------|-------------------------------|--------------------------|
| Al <sub>2</sub> O <sub>3</sub> | 0.712                     | 0.402                    | 0.790                     | 0.448                         | 0.692                    |
| Cr <sub>2</sub> O <sub>3</sub> | 0.102                     | 0.044                    | 0.098                     | 0.212                         | 0.080                    |
| CuO                            | 0.007                     | 0.003                    | 0.003                     | 0.002                         | 0.004                    |
| FeO                            | 43.956                    | 44.589                   | 45.195                    | 55.362                        | 51.786                   |
| MgO                            | 0.116                     | 0.084                    | 0.218                     | 0.402                         | 0.370                    |
| MnO                            | 1.517                     | 1.443                    | 1.495                     | 1.072                         | 0.801                    |
| NiO                            | 0.007                     | 0.006                    | 0.006                     | 0.004                         | 0.000                    |
| TiO <sub>2</sub>               | 28.086                    | 35.093                   | 42.638                    | 36.879                        | 40.810                   |
| V <sub>2</sub> O <sub>3</sub>  | 0.257                     | 0.245                    | 0.409                     | 0.277                         | 0.243                    |
| ZnO                            | 0.063                     | 0.073                    | 0.056                     | 0.031                         | 0.032                    |
| Totals                         | 74.839                    | 81.982                   | 90.908                    | 94.689                        | 94.818                   |

(b) Shelf areas

|                                | Long<br>Island<br>Shelf<br>(n=11) | Atlantic<br>Shelf<br>area 1<br>(n=37) | Atlantic<br>Shelf<br>area 2<br>(n=45) | Atlantic<br>Shelf<br>area 3<br>(n=24) | Atlantic<br>Shelf<br>area 4<br>(n=8) |
|--------------------------------|-----------------------------------|---------------------------------------|---------------------------------------|---------------------------------------|--------------------------------------|
| Al <sub>2</sub> O <sub>3</sub> | 1.035                             | 1.021                                 | 0.839                                 | 0.661                                 | 0.802                                |
| Cr <sub>2</sub> O <sub>3</sub> | 0.096                             | 0.078                                 | 0.098                                 | 0.129                                 | 0.136                                |
| CuO                            | 0.004                             | 0.004                                 | 0.004                                 | 0.003                                 | 0.003                                |
| FeO                            | 64.908                            | 42.260                                | 44.108                                | 45.918                                | 48.753                               |
| MgO                            | 0.221                             | 0.351                                 | 0.393                                 | 0.343                                 | 0.515                                |
| MnO                            | 1.177                             | 1.207                                 | 1.107                                 | 1.238                                 | 1.000                                |
| NiO                            | 0.032                             | 0.010                                 | 0.009                                 | 0.006                                 | 0.027                                |
| TiO <sub>2</sub>               | 18.869                            | 41.355                                | 43.556                                | 43.345                                | 37.545                               |
| V <sub>2</sub> O <sub>3</sub>  | 0.230                             | 0.180                                 | 0.205                                 | 0.197                                 | 0.193                                |
| ZnO                            | 0.025                             | 0.041                                 | 0.042                                 | 0.044                                 | 0.042                                |
| Totals                         | 86.597                            | 86.507                                | 90.361                                | 91.884                                | 89.016                               |

About 0.1 g Fe-Ti oxide grains were separated from the 0.063–0.25 mm fraction by a Frantz magnetic separator (0.1–0.3 A), fused with K-pyrosulphate, dissolved in nitric acid and analysed by flame atomic absorption<sup>3</sup>. All Fe was assigned to FeO rather than assuming stoichiometric proportions of FeO and Fe<sub>2</sub>O<sub>3</sub> for ilmenite because some haematite and titanomagnetite were always present. The low totals reflect primarily the oxygen that would be combined as Fe<sub>2</sub>O<sub>3</sub>. All statistical analyses were carried out on element rather than oxide content. All contents are in weight per cent.



TABLE 2 Results of discriminant function analysis

| Atlantic Shelf area                            | % Hudson River | % Long Island Shelf | % Susquehanna River | % James River and others | % Hudson River and Long Island Shelf | % Other Combinations involving Hudson or Long Island | Estimated total Hudson and Long Island contribution |
|------------------------------------------------|----------------|---------------------|---------------------|--------------------------|--------------------------------------|------------------------------------------------------|-----------------------------------------------------|
| 1, Long Island to Cape May<br>( <i>n</i> =37)  | 41<br>(8)      | 24<br>(24)          | 0                   | 5<br>(5)                 | 0                                    | 30                                                   | 70–85%                                              |
| 2, Cape May to Cape Charles<br>( <i>n</i> =45) | 67<br>(18)     | 4<br>(2)            | 4<br>(2)            | 0                        | 7                                    | 18                                                   | 70–80%                                              |
| 3, Cape Charles to Hatteras<br>( <i>n</i> =24) | 54<br>(8)      | 4<br>(4)            | 21<br>(4)           | 8<br>(0)                 | 0                                    | 8                                                    | 60–70%                                              |
| 4, Hatteras to Cape Fear<br>( <i>n</i> =8)     | 25<br>(25)     | 25<br>(25)          | 13<br>(0)           | 0                        | 25                                   | 13                                                   | 50–60%                                              |

These results are the percentages of shelf samples classified with potential sources (probability >0.95). The percentage of shelf samples from each area that have a significant contribution of grains from an unsampled source is given in parentheses. The probability that the discriminant function score of these samples would plot this far from the centroid of the group to which they were classified was <0.5 (ref. 24). For example, all of the samples classified into the Long Island and James River source groups from area 1 most likely came in part or entirely from unsampled sources. The use of this probability avoids forced classifications where shelf samples are assigned to a source group because their Fe–Ti oxide composition is more similar to one source than any other sampled source.

uncertainty about sources in this part of the shelf and elsewhere can probably be accounted for by local sources, such as shore erosion, or complex mixtures of sand with no single dominant source.

Southward movement of sand has long been known to occur on this shelf, but the timing and processes of sand movement between shelf compartments during the Pleistocene are uncertain<sup>14–16</sup>. Textural patterns of shelf sediment suggest that insufficient time has elapsed since the last transgression (~17,000 years ago) for storm currents to winnow out the relatively finer sand and transport it beyond a single shelf compartment (Fig. 1)<sup>17</sup>. The more efficient transport of sand in the near-shore zone by longshore drift is also restricted to a single compartment during a given transgression because as sand moves southward it becomes trapped in the next large bay-mouth shoal or estuary until the sea level drops<sup>18</sup>. During the next drop in the sea level, streams would flush the trapped sand onto the shelf to be transported farther south by net littoral drift and storm currents beyond the nearshore area. As levels drop further, rejuvenated rivers reoccupy their shelf channels and a series of thin deltas form from which the sand is partially removed by fluvial erosion<sup>15,19</sup>. Much of this eroded sand could be lost to canyons especially during the low stand. During the subsequent landward transgression some of the sand stored in the dissected deltas would be reworked by shoreface erosion and transported south by the littoral drift system to the next estuary to the south. Thus several sea-level cycles would be required to transport Hudson River sand to the North Carolina shelf (area 3, Fig. 1).

With each cycle some of the less resistant minerals are removed from coastal plain soils by weathering. When these weathered grains are eroded and/or flushed out of estuaries by fluvial processes during a period of decreasing sea level, characteristic abundance patterns develop across the shelf<sup>1,7</sup>. Patterns of the quartz-grain shapes across the shelf are similarly affected because these patterns are due to grains rounded by weathering<sup>20</sup>. Despite the superposition of abundance or shape patterns, the Fe–Ti oxide picture is unaffected. Only relatively unweathered Fe–Ti oxide grains were analysed in this study and their compositions is that of their primary source rocks. Local contributions of fresh Fe–Ti oxide grains apparently have only a small effect on the composition of shelf samples as indicated by the decrease in northern source associations from shelf area 2 to 3 (Table 2).

The dominance of sand from the Hudson River and other New England sources in shelf sand is due to the drainage of glacial meltwater especially during the initial phases of a sea level rise. These sources contribute little to the shelf today.

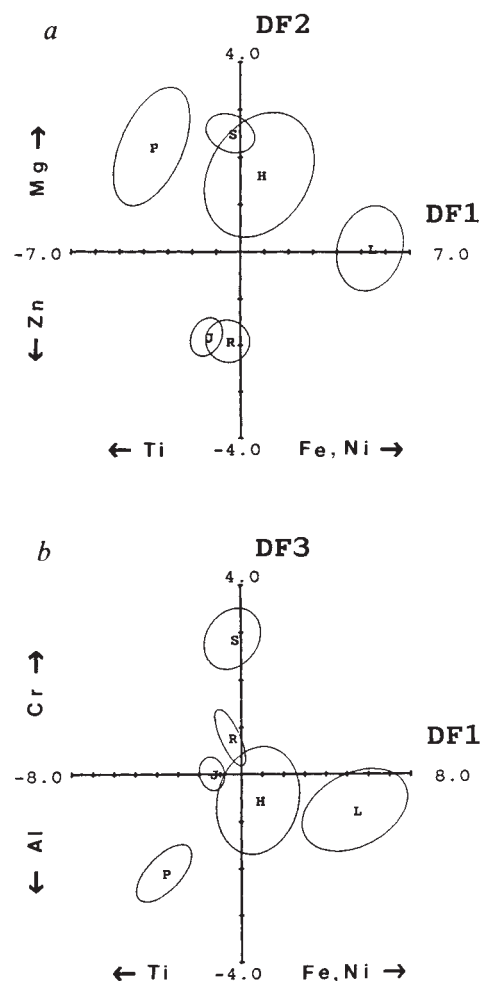


FIG. 2 Probability (0.95) ellipsoids about source-group centroids for *a*, discriminant function 1 (DF1) plotted against DF2 and *b*, DF1 plotted against DF3. The separation among groups is good. DF1 accounts for 54% of the variance, DF2 for 24%, and DF3 for 13%. The remaining variance (9%) is accounted for by DF4 and DF5, not plotted. The contribution by the most important variables is shown by arrows. Thus whereas Cr and Al are most important in distinguishing the Hudson River (H) and Susquehanna River (S), Ti, Fe and Ni distinguish the Hudson River and Long Island shelf group (L). The Potomac River (P), Roanoke River (R) and James River (J) are most distinguished by Mg, Cr, Zn and Al.

Recent discoveries of buried sub-glacially-cut tunnels on the Scotian Shelf<sup>21</sup> and similar ideas on the origin of the New York Finger Lakes<sup>22</sup> suggest that huge floods may have occurred during melting of the Laurentide ice sheet<sup>23</sup>. These floods could flush large volumes of sand onto the then sub-aerially exposed shelf.

The Hatteras promontory obviously had a major role in funnelling sand off this shelf and into deeper water especially during times of lower sea level. Only south of this promontory does the Hudson River and Long Island shelf Fe-Ti oxide signature diminish substantially. □

Received 29 January; accepted 11 July 1990.

1. Milliman, J. D., Pilkey, O. H. & Ross, D. A. *Geol. Soc. Am. Bull.* **83**, 1315-1334 (1972).
2. McMaster, R. L. & Garrison, L. E. *J. sedim. Petrol.* **36**, 1131-1142 (1966).
3. Darby, D. A. & Tsang, Y. W. *J. sedim. Petrol.* **57**, 831-838 (1987).
4. Morton, A. C. in *Provenance of Arenites* 249-277 (Reidel, Boston, 1985).
5. Rumble, D. III *Oxide Minerals* (Min. Soc. Am., Washington, DC, 1976).
6. Kelling, G., Sheng, H. & Stanley, D. J. *Geol. Soc. Am. Bull.* **86**, 853-862 (1975).
7. Ross, D. A. *Prof. Pap. U.S. geol. Surv.* 529-G, 40 (1970).
8. Grosz, A. E. *Mar. Min.* **6**, 339-357 (1987).
9. Darby, D. A. *Geol. Soc. Am. Bull.* **95**, 1208-1218 (1984).
10. Buddington, A. F. & Lindsley, D. H. *J. Petrol.* **5**, 310-357 (1964).
11. Ramdohr, P. *Ore Minerals and Their Intergrowths* (Pergamon, New York, 1980).
12. Knebel, H. J. *Mar. Geol.* **42**, 349-368 (1981).
13. Puffer, J. H. & Cousminer, H. L. *Econ. Geol.* **77**, 379-391 (1982).
14. Duane, D. B., Field, M. E., Meisburger, E. P., Swift, D. J. P. & Williams, S. J. in *Shelf Sediment Transport*, 447-498 (Dowden, Hutchinson & Ross, Stroudsburg, 1972).
15. Swift, D. J. P., Kofoed, J. W., Saulsbury, F. P. & Sears, P. in *Shelf Sediment Transport*, 499-574 (Dowden, Hutchinson & Ross, Stroudsburg, 1972).
16. Vincent, C. E., Swift, D. J. P. & Hillard, B. *Mar. Geol.* **42**, 369-398 (1981).
17. Swift, D. J. P., Thorne, J. A. & Oertel, G. F. in *Shelf Sands and Sandstones*, 191-211 (Can. Soc. Petrol. Geol., Calgary, 1986).
18. Swift, D. J. P. in *Marine Sediment Transport and Environmental Management*, 255-310 (Wiley, New York, 1976).
19. Swift, D. J. P. *Geol. Soc. Am. Bull.* **84**, 2743-2748 (1973).
20. Prusak, D. & Mazzullo, J. *J. sedim. Petrol.* **57**, 278-287 (1987).
21. Boyd, R., Scott, D. B. & Douma, M. *Nature* **333**, 61-64 (1988).
22. Mullins, H. T. & Hinchey, E. J. *Geology* **17**, 622-625 (1989).
23. Shaw, J. *Geology* **17**, 853-856 (1989).
24. Klecka, W. R. in *Statistical Package for the Social Sciences*, 434-467 (McGraw-Hill, Toronto, 1975).

ACKNOWLEDGEMENTS. I thank D. J. P. Swift, A. E. Grosz and G. R. Whittecar for suggestions. The US Geological Survey supplied most of the shelf samples and Virginia Institute of Marine Sciences supplied the Chesapeake Bay and Bay Mouth samples. I thank Y. W. Tsang, L. Tsang, E. A. Council and M. Lawless for assistance in sample collection and analysis.

## U-Pb evidence for Abitibi gold mineralization postdating greenstone magmatism and metamorphism

R. A. Jemielita<sup>†</sup>, D. W. Davis<sup>\*</sup> & T. E. Krogh<sup>\*</sup>

<sup>\*</sup> Jack Satterly Geochronology Laboratory, Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario, Canada, M5S 2C5

<sup>†</sup> Department of Geology, University of Toronto, 22 Russell Street, Toronto, Ontario, Canada, M5S 3B1

TWO widely held theories for the origin of Archaean lode gold deposits, the magmatic<sup>1</sup> and metamorphic replacement<sup>2</sup> models, predict that gold emplacement should have occurred near the end of late, syn-kinematic greenstone magmatism. Here we report that at the Camflo mine in the Abitibi subprovince of Canada, U-Pb ages for rutile and titanite in auriferous alteration ( $2,625 \pm 7$  Myr) are much younger than the age of the host quartz syenite ( $2,680 \pm 4$  Myr) and ages for regional magmatism ( $>2,670$  Myr)<sup>3</sup> and metamorphism ( $\sim 2,680$  Myr)<sup>4</sup>. Mineral associations and Pb-Pb systematics indicate that the hydrothermal minerals record crystallization ages that are probably coeval with the gold mineralization. These ages, and others from the Abitibi Subprovince<sup>4-6</sup>, rule out a direct connection between gold and magmatism or metamorphism associated with regional folding of the greenstone belt. They are broadly contemporaneous with metamorphism at deep crustal levels<sup>7</sup>, and may record late devolatilization of the lower crust.

Archaean greenstone belts have produced  $\sim 14,000$  tonnes of gold, the origin of which is of great geological and economic interest<sup>8</sup>. The southern Abitibi subprovince in the Superior province is one of the world's richest gold-producing areas. Deposits are concentrated along two structural breaks, the Destor-Porcupine and Cadillac Faults<sup>9</sup>. U-Pb zircon data from this region<sup>3</sup> show that most volcanism ended at  $\sim 2,700$  Myr and was followed by folding and intrusion of late plutons. Magmatic activity ended by  $\sim 2,670$  Myr.

The earliest attempts to bracket the age of gold in this area involved  $^{39}\text{Ar}$ - $^{40}\text{Ar}$  dating of hydrothermal minerals. Fuchsite from Dome mine gave an age of  $2,633 \pm 6$  Myr<sup>5</sup>, whereas muscovite from Sigma mine gave an age of  $2,579 \pm 3$  Myr and was claimed to date gold emplacement<sup>10</sup>. These results were not widely accepted as dating the mineralization because of the possibility that they might represent cooling ages. They have been supported by similar U-Pb ages in ore-associated rutile<sup>4,6</sup>. The thermal closure temperature for diffusion of Pb in this mineral is  $\sim 400^\circ\text{C}$ <sup>11</sup>.

The Camflo mine was chosen for dating because the gold-related alteration there contains titanite and rutile<sup>12</sup>, and titanite is known to have a blocking temperature above temperatures of regional metamorphism (mid-to-upper greenschist facies)<sup>11</sup>. The mine is located in the Malartic-Val d'Or area,  $\sim 3$  km north of the Cadillac break. By 1986 it had produced 51 tonnes of gold, mainly from disseminated ore and quartz veins in a small ( $\sim 200$ -m diameter) irregular stock of porphyritic quartz syenite that intrudes folded metavolcanic and metasedimentary rocks.

Only a few unaltered zircon grains were obtained from a sample of unmineralized quartz syenite, which also contained abundant, pale-brown titanite. The mineralized quartz syenite (ore zone) sample contained little titanite but had abundant, mostly dark rutile. Three quartz-carbonate vein samples contained euhedral titanite crystals 2-3 mm long. A fourth, auriferous, vein sample contained scheelite, abundant black rutile, and a small amount of titanite. U-Pb isotope analyses were carried out on fractions from each of these mineral separates.

Sample processing, analytical and statistical techniques followed conventional procedures<sup>13-15</sup>. Age errors and error ellipses (Figs 1-3) are quoted at the 95% confidence level. Initial Pb isotopic compositions for titanite in the unmineralized quartz

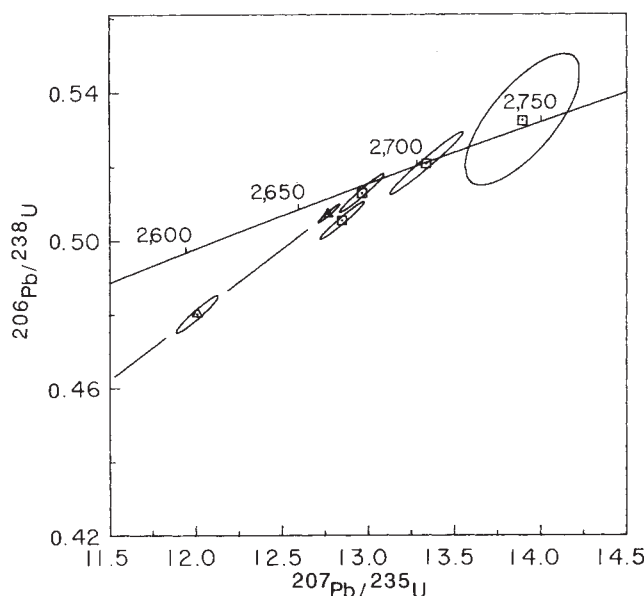


FIG. 1 U-Pb data for zircon (squares) and titanite (triangles) from an unmineralized sample of the Camflo stock quartz syenite. The regression line shown is for data points from analyses of two titanite fractions. The zircons contain an inherited component.



TABLE 1 U-Pb isotope data

| Zircon fractions                                     | Weight (mg) | U (p.p.m.) | Common Pb* (pg) | $\frac{^{207}\text{Pb}}{^{206}\text{Pb} \dagger}$ | $\frac{^{207}\text{Pb}}{^{204}\text{Pb} \dagger}$ | $\frac{^{208}\text{Pb}}{^{206}\text{Pb} \dagger}$ | $\frac{^{206}\text{Pb}}{^{238}\text{U} \ddagger}$ | $\frac{^{207}\text{Pb}}{^{235}\text{U} \ddagger}$ | $\frac{^{207}\text{Pb}}{^{206}\text{Pb} \ddagger \S}$<br>Age (Myr) |
|------------------------------------------------------|-------------|------------|-----------------|---------------------------------------------------|---------------------------------------------------|---------------------------------------------------|---------------------------------------------------|---------------------------------------------------|--------------------------------------------------------------------|
| (a) Unmineralized quartz syenite                     |             |            |                 |                                                   |                                                   |                                                   |                                                   |                                                   |                                                                    |
| Titanite Ab                                          | 0.940       | 27         | 1,784           | 0.20781                                           | 100.8<br>(0.4)                                    | 0.159                                             | 0.50729<br>(0.50)                                 | 12.773<br>(0.10)                                  | 2,676.7<br>(1.4)                                                   |
| Titanite Ab                                          | 0.227       | 22         | 543             | 0.22248                                           | 65.2<br>(1.0)                                     | 0.115                                             | 0.48007<br>(1.00)                                 | 12.008<br>(0.20)                                  | 2,665.7<br>(6.3)                                                   |
| Zircon Ab<br>1 grain                                 | 0.001       | 92         | 4               | 0.18334                                           |                                                   | 0.114                                             | 0.51292<br>(1.00)                                 | 12.966<br>(0.20)                                  | 2,683.2<br>(0.6)                                                   |
| Zircon Ab<br>1 grain                                 | 0.003       | 71         | 5               | 0.18440                                           |                                                   | 0.195                                             | 0.50539<br>(1.00)                                 | 12.850<br>(0.20)                                  | 2,693.0<br>(2.5)                                                   |
| Zircon Ab<br>1 grain                                 | 0.003       | 50         | 3               | 0.18575                                           |                                                   | 0.194                                             | 0.52089<br>(1.60)                                 | 13.340<br>(0.20)                                  | 2,705.0<br>(0.1)                                                   |
| Zircon Ab<br>2 fragments                             | 0.001       | 11         | 4               | 0.18930                                           |                                                   | 0.059                                             | 0.53245<br>(2.40)                                 | 13.898<br>(1.00)                                  | 2,736.0<br>(0.7)                                                   |
| (b) Mineralized quartz syenite                       |             |            |                 |                                                   |                                                   |                                                   |                                                   |                                                   |                                                                    |
| Titanite Ab                                          | 0.137       | 30         | 861             | 0.25135                                           | 41.3<br>(0.6)                                     | 0.157                                             | 0.49439<br>(1.00)                                 | 12.082<br>(0.27)                                  | 2,627.2<br>(1.7)                                                   |
| Rutile Ab<br>Dark                                    | 0.494       | 8          | 409             | 0.21404                                           | 70.7<br>(0.4)                                     | 0.251                                             | 0.49990<br>(0.80)                                 | 12.208<br>(0.13)                                  | 2,626.0<br>(0.6)                                                   |
| Rutile Ab<br>Pale brown                              | 0.045       | 8          | 57              | 0.22900                                           | 54.6<br>(0.5)                                     | 0.145                                             | 0.48462<br>(0.81)                                 | 11.888<br>(0.26)                                  | 2,633.4<br>(4.0)                                                   |
| (c) Quartz-titanite veins                            |             |            |                 |                                                   |                                                   |                                                   |                                                   |                                                   |                                                                    |
| Titanite Ab                                          | 0.397       | 24         | 2,283           | 0.26109                                           | 38.0<br>(0.5)                                     | 0.143                                             | 0.50134<br>(0.81)                                 | 12.260<br>(0.26)                                  | 2,628.2<br>(0.4)                                                   |
| Titanite Ab                                          | 0.528       | 23         | 2,641           | 0.25153                                           | 41.3<br>(0.4)                                     | 0.139                                             | 0.49990<br>(1.01)                                 | 12.220<br>(0.20)                                  | 2,627.5<br>(0.7)                                                   |
| Titanite Ab                                          | 0.127       | 19         | 875             | 0.29567                                           | 30.4<br>(0.5)                                     | —                                                 | 0.49347<br>(1.04)                                 | 12.047<br>(0.39)                                  | 2,625.4<br>(1.8)                                                   |
| (d) Auriferous quartz-scheelite-rutile-titanite vein |             |            |                 |                                                   |                                                   |                                                   |                                                   |                                                   |                                                                    |
| Rutile Ab<br>Black                                   | 0.399       | 11         | 1,690           | 0.29626                                           | 29.7<br>(0.2)                                     | 0.221                                             | 0.49860<br>(1.0)                                  | 11.983<br>(0.2)                                   | 2,599.3<br>(-0.4)                                                  |
| Rutile Ab<br>Black                                   | 1.202       | 9          | 3,505           | 0.27913                                           | 32.7<br>(1.0)                                     | 0.207                                             | 0.49928<br>(1.0)                                  | 12.031<br>(0.1)                                   | 2,603.7<br>(-0.3)                                                  |
| Titanite Ab                                          | 0.081       | 34         | 495             | 0.24494                                           | 43.8<br>(0.4)                                     | 0.139                                             | 0.45907<br>(2.0)                                  | 11.192<br>(0.2)                                   | 2,623.2<br>(8.6)                                                   |
| Scheelite Ab                                         | 2.975       | 0.028      | 102,727         | 1.08039                                           | 14.580<br>(0.30)                                  | 2.468                                             | —                                                 | —<br>(0.1)                                        | —                                                                  |
| Scheelite                                            | 5.470       | 0.033      | 192,653         | 1.07982                                           | 14.582<br>(0.20)                                  | 2.465                                             | —                                                 | —<br>(0.1)                                        | —                                                                  |

Ab, abraded fraction. Numbers in brackets in columns 6, 8 and 9 are  $2\sigma$  errors (%); in the last column the number in brackets is the discordance (%). Errors in column 8 are for Pb/U; those in column 9 are for the  $^{207}\text{Pb}/^{206}\text{Pb}$  age (ratio, in the case of the scheelite analyses).

\* Includes blank and initial common lead.

† Corrected for fractionation, blank and spike lead. U blank = 3.8 pg (rutile, titanite and scheelite), 0.5 pg (zircon). Pb blank < 8 pg. Pb and U fractionation correction = 0.13% per a.m.u.

‡ Corrected for fractionation, blank and common Pb.

§  $^{238}\text{U}$  decay constant =  $0.15513 \times 10^{-9} \text{ yr}^{-1}$ .  $^{235}\text{U}$  decay constant =  $0.98485 \times 10^{-9} \text{ yr}^{-1}$ .  $^{238}\text{U}/^{235}\text{U} = 137.88$ .

|| Common lead content is at the blank level.

syenite are assumed to lie on the Stacey and Kramers<sup>16</sup> growth curve. For titanite and rutile in the mineralized quartz syenite and the veins, common Pb is assumed to have the same initial isotopic composition as that measured in the scheelite ( $^{207}\text{Pb}/^{204}\text{Pb} = 14.581$ ,  $^{206}\text{Pb}/^{204}\text{Pb} = 13.498$ ). This value can be accurately determined for scheelite because of its low U/Pb ratio ( $< 10^{-3}$ ) and because the scheelite is chemically inert and unaltered.

The data for four uncracked, unaltered zircon fractions from the unmineralized quartz syenite show an older component due to inheritance (Fig. 1). The  $^{207}\text{Pb}/^{206}\text{Pb}$  age of the youngest concordant zircon fraction (analysis 3, Table 1) provides an upper age limit for crystallization of the stock ( $2,683 \pm 3$  Myr). Because titanite is not susceptible to magmatic inheritance, the most concordant titanite fraction gives a minimum age for the stock ( $2,677 \pm 2$  Myr). Regression of the two titanite points gives an upper intercept age of  $2,680 \pm 4$  Myr (Fig. 1). As this overlaps the  $^{207}\text{Pb}/^{206}\text{Pb}$  ages of both the oldest titanite and the youngest

zircon fractions, it is a reliable age estimate for intrusion of the Camflo stock.

Data points for titanite fractions from the veins and ore zone overlap and, together with a discordant titanite fraction, define an age of  $2,627 \pm 2$  Myr (Fig. 2). The age of a dark rutile fraction, representative of the main rutile population from the mineralized quartz syenite, is in agreement with this but a subordinate pale-brown fraction from the same rock is somewhat discordant and slightly older ( $2,633 \pm 3$  Myr). The rutile fractions from the scheelite-bearing vein are concordant, but distinctly younger ( $2,600 \pm 3$  Myr).

Magmatic titanite in the quartz syenite probably reacted to form rutile in the ore zone, or recrystallized, completely resetting the U-Pb system. Resetting was perhaps incomplete for the pale-brown rutile fraction (analysis 9). Alternatively, this fraction may have had a different initial common Pb composition.

Data points derived from the six analyses that have similar ages on the concordia diagram are collinear on a plot of

$^{207}\text{Pb}/^{204}\text{Pb}$  against  $^{206}\text{Pb}/^{204}\text{Pb}$  (75% probability of fit) and give an age of  $2,625 \pm 7$  Myr (Fig. 3), which is independent of assumptions about the isotopic composition of initial common Pb. These points give a much better fit with the scheelite Pb isotope value than the Stacey and Kramers value (50% and 11% probability of fit, respectively), indicating that these mineral populations may have crystallized from the same fluid. Inclusion of either the older or younger rutile data points reduces the probability of fit almost to zero, therefore some rutile populations have different ages and/or initial Pb isotopic compositions.

The event at  $2,625 \pm 7$  Myr is probably coeval with gold because: (1) abundant rutile and the younger titanite populations were found only in the ore-grade material, not in the unmineralized sample; (2) titanite and rutile have greatly different thermal closure temperatures for diffusion of lead<sup>11</sup>, so their age agreement indicates that they are undisturbed; (3) tungsten shows a positive correlation with gold<sup>17</sup> and, because scheelite gives a lead isotopic composition collinear, within error, with the  $2,625 \pm 7$  Myr old hydrothermal minerals, these minerals and gold probably crystallized from the same fluid; (4) a field study of the Camflo mine<sup>12</sup> suggests that these hydrothermal minerals are associated with or pre-date the gold. Isotope data obtained by Zweng and Mortensen<sup>18</sup> in an independent study of these minerals give similar age results.

Gold-related alteration at Camflo mine and elsewhere in the Malartic-Val d'Or region has been previously shown to overprint regional greenschist-facies metamorphism<sup>12,19</sup>. This is dated at  $2,680 \pm 2$  Myr by U-Pb analyses of metamorphic rutile<sup>4</sup> and  $2.69\text{--}2.67$  Gyr by  $^{39}\text{Ar}/^{40}\text{Ar}$  analyses of metamorphic hornblende<sup>10</sup>. The alteration at Camflo therefore significantly post-dates metamorphism as well as all Archaean magmatism known in the Abitibi Subprovince<sup>3</sup>. These data are in accord with results from other mines, mentioned above, which indicate very late ages for gold mineralization<sup>4-6</sup>. The results of this study are also similar to U-Pb zircon and rutile ages that constrain the timing of some gold-related alteration at Hemlo<sup>20</sup> and in the Yilgarn Block, Australia<sup>21,22</sup>.

The only chronologically documented events in the Superior Province that occur this late are growth of zircon in high-grade metamorphic rocks of the Kapuskasing structural zone ( $2,660\text{--}2,584$  Myr)<sup>7,23</sup>. This structure may represent an uplifted cross-

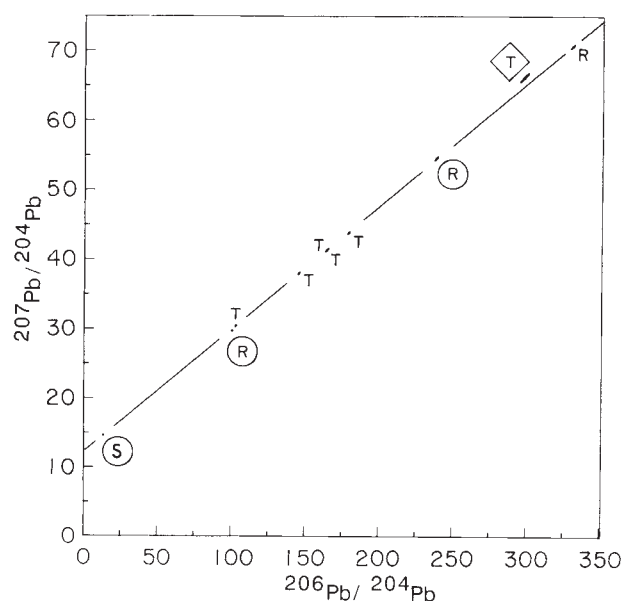


FIG. 3 Pb-Pb data for titanite (T), rutile (R) and scheelite (S) fractions. The line and age are determined from a best fit to points represented by the unenclosed symbols (titanite from the quartz-carbonate veins, and titanite and rutile from the mineralized quartz syenite). The circled symbols represent scheelite and younger rutile fractions from one auriferous, scheelite-bearing vein. The diamond symbol represents an older titanite fraction from the unmineralized quartz syenite.

section through the Archaean crust<sup>24</sup> and thus be representative of rocks underlying the greenstone belts. If so, the age correspondence between metamorphism at deep crustal levels and hydrothermal activity at shallower levels suggests that they are genetically related, as previously noted<sup>10</sup>. Fluids derived from devolatilization of the lower crust and focused along structural breaks (such as the Cadillac break) may have been involved in late hydrothermal activity and gold mineralization.

Although these data rule out high-level magmatism and metamorphism as a source for some gold deposits, it has been shown that gold mineralization in the Red Lake and Wabigoon greenstone belts, in the central Superior Province, was coeval with regional magmatism and metamorphism<sup>25,26</sup>. Archaean lode gold mineralization may therefore result from different processes occurring at different times in the evolution of greenstone belts. The very late, and presumably deep, thermal processes responsible for the large Abitibi deposits may also be an integral part of the development of the greenstone belt. □

**Editorial note:** The reference in this Letter's acknowledgements to mapping by P. L. Zweng requires some explanation. The collection of the samples dated here depended on geological maps compiled by Paul Zweng for his own graduate research at Stanford University (see ref. 18 above). These maps were acquired at the Camflo mine by a third party, then an associate of Jemielita's, who did not have Zweng's permission to use the maps for this purpose. Although the Editor regrets the circumstances and the contention they have caused, he believes that the interest of the data requires that publication should not be denied or delayed. □

Received 17 January; accepted 12 July 1990.

- Burrows, D. R. & Spooner, E. T. C. *Econ. Geol. Monogr.* **6**, 424-444 (1988).
- Groves, D. I., Phillips, N., Ho, S. E., Houston, S. M. & Standing, C. A. *Econ. Geol.* **82**, 2045-2058 (1987).
- Corfu, F., Krogh, T. E., Kwok, K. & Jensen, L. S. *Can. J. Earth Sci.* **26**, 1747-1763 (1989).
- Wong, L. et al. *Geol. Ass. Can. Miner. Ass. Can. Mtg.* A45 (1989).
- Masilvec, A., McMaster, D. & York, D. in *Ontario Geol. Surv. Misc. Pap.* **130**, 107-114 (1986).
- Jemielita, R. A., Davis, D. W., Krogh, T. E. & Spooner, E. T. C. *Geol. Soc. Am. Abstr. Progr.* A351 (1989).
- Krogh, T. E., Heaman, L. M. & Machado, N. T. *Kapuskasing Structural Zone Transect Project, Lithoprobe Workshop*, 243 (University of Toronto, 1988).

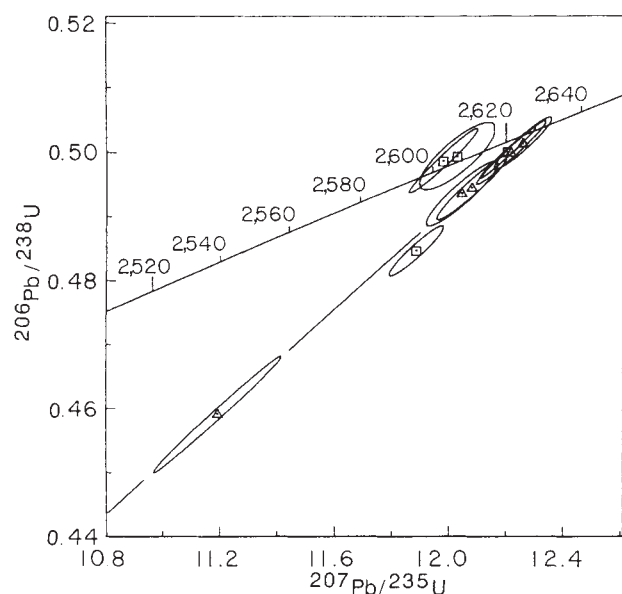


FIG. 2 U-Pb data for hydrothermal titanite (triangles) and rutile (squares) fractions from mineralized (ore zone) Camflo stock quartz syenite and quartz-carbonate veins. The regression line represents a best fit for data points from hydrothermal titanite in four veins and the mineralized quartz syenite, as well as hydrothermal rutile from the latter.



8. Roberts, R. G. *Geosci. Can.* **76**, 37–52 (1987).
9. Latulippe, M. *Can. Inst. Min. Metal. Spec. Vol.* **24**, 9–14 (1982).
10. Hanes, J. A., Archibald, D. A., Hodgson, C. J. & Robert, F. *Geol. Surv. Can. Pap.* **89-1C**, 135–142 (1989).
11. Mezger, K., Hanson, G. N. & Bohlen, S. R. *Earth planet. Sci. Lett.* **96**, 106–118 (1989).
12. Sauve, P. & Makila, A. *Ministère de l'Énergie et des Ressources du Québec, Rapport MB 87-40* (1987).
13. Krogh, T. E. *Geochim. cosmochim. Acta* **37**, 485–494 (1973).
14. Krogh, T. E. *Geochim. cosmochim. Acta* **46**, 637–649 (1982).
15. Davis, D. W. *Can. J. Earth Sci.* **19**, 2141–2149 (1982).
16. Stacey, J. S. & Kramers, J. D. *Earth planet. Sci. Lett.* **34**, 207–226 (1975).
17. Chainey, D. thesis, Ecole Polytechnique de Montreal (1983).
18. Zweng, P. L. & Mortensen, J. K. *Geol. Soc. Am. Abstr. Progr.* A351 (1989).
19. Robert, F. & Brown, A. C. *Econ. geol.* **81**, 593–616 (1986).
20. Corfu, F. & Muir, T. L. *Chem. Geol.* **79**, 201–223 (1989).
21. Clarke, M. E., Carmichael, D. M., Hodgson, C. J. & Fu, M. *Econ. Geol. Monogr.* **6**, 445–459 (1988).
22. Hill, R. I. & Campbell, I. H. *Aust. J. Earth Sci.* **36**, 313–316 (1989).
23. Corfu, F. *Precambrian Res.* **36**, 259–275 (1987).
24. Percival, J. A. & Card, K. D. *Geology* **11**, 323–326 (1983).
25. Corfu, F. & Andrews, A. J. *Can. J. Earth Sci.* **23**, 107–109 (1986).
26. Davis, D. W. & Smith, P. M. *J. Geol.* (in the press).

ACKNOWLEDGEMENTS. Mapping carried out by P. L. Zweng was used in the collection of samples. We thank the management at the Camflo mine and American Barrick Resources Corporation for access to their property, in particular A. Makila, J. Daigneault, N. Lacroix and R. Bergeron. We also thank P. Sauve and D. I. Groves for reviews. R. A. J. acknowledges the financial support of the Natural Sciences and Engineering Research Council of Canada through grants awarded to E. T. C. Spooner.

## A new look at differentiation of the Earth from melting experiments on the Allende meteorite

Carl B. Agee\*

Bayerisches Geoinstitut, Universität Bayreuth, Postfach 101251, 8580 Bayreuth, FRG

**CARBONACEOUS** chondrites contain approximately solar abundances of the non-volatile elements<sup>1</sup>, but are much more FeO-rich<sup>2</sup> than upper mantle peridotites<sup>3</sup> or hypothetical 'chondritic' mantle compositions<sup>4,5</sup>. If the Earth accreted from unfractionated, primitive meteoritic debris similar to the carbonaceous chondrites, how then did it become a fractionated, layered body with a crust, mantle and core? Here I report the results of high-pressure melting experiments on the Allende CV3 carbonaceous chondrite, which address this question and provide a new look at the Earth's earliest stage of differentiation. Multi-anvil experiments at 24, 26 and 26.5 GPa show that FeO-rich magnesiowüstite is an abundant crystallizing phase at temperatures near the Allende silicate liquidus. If a chondritic Earth experienced a high-temperature molten stage, then during cooling and crystallization, FeO-rich magnesiowüstite could be segregated to the deepest levels of the Earth's interior. Magnesiowüstite fractionation may thus have depleted the initial FeO content of the primitive chondritic mantle and contributed to the formation and growth of the Earth's core.

Experiments were performed using a split-sphere multi-anvil device. Magnesiowüstite and majorite garnet were identified as co-crystallizing liquidus phases in Allende at 24 GPa. The silicate solidus at this pressure contains the assemblage magnesiowüstite + majorite +  $\gamma$ -spinel. At 26 and 26.5 GPa magnesiowüstite and majorite remain the first liquidus phases, but when the temperature is decreased slightly ( $\sim 10^\circ\text{C}$ ), majorite is apparently consumed in a reaction that crystallizes Mg-perovskite + magnesiowüstite in the presence of silicate liquid. The silicate solidus phase assemblage in Allende at 26 GPa is Mg-perovskite + magnesiowüstite, with minor ( $< 1\%$ ) Ca-perovskite observed as small ( $< 2\ \mu\text{m}$ ) inclusions in Mg-perovskite. Electron-microprobe analyses of crystalline phases and coexisting melt from an experiment at 26 GPa are given in Table 1. Tem-

perature at the liquidus is estimated to be  $2,050 \pm 50^\circ\text{C}$ . Figure 1 consists of scanning electron micrographs from different regions in the experimental charge where magnesiowüstite, majorite, perovskite and quench melt are observed. Coexisting with the high-pressure-temperature silicate/oxide assemblage is interstitial sulphide quench melt with microprobe-determined element proportions  $\text{Fe}_{0.41}\text{Ni}_{0.10}\text{Re}_{0.01}\text{S}_{0.48}$ . Chemical analyses<sup>2</sup> show that the sulphide component (pentlandite) comprises  $\sim 6$ –7 wt% of the bulk Allende meteorite, with Fe and Ni metal accounting for another  $\sim 1$  wt%.

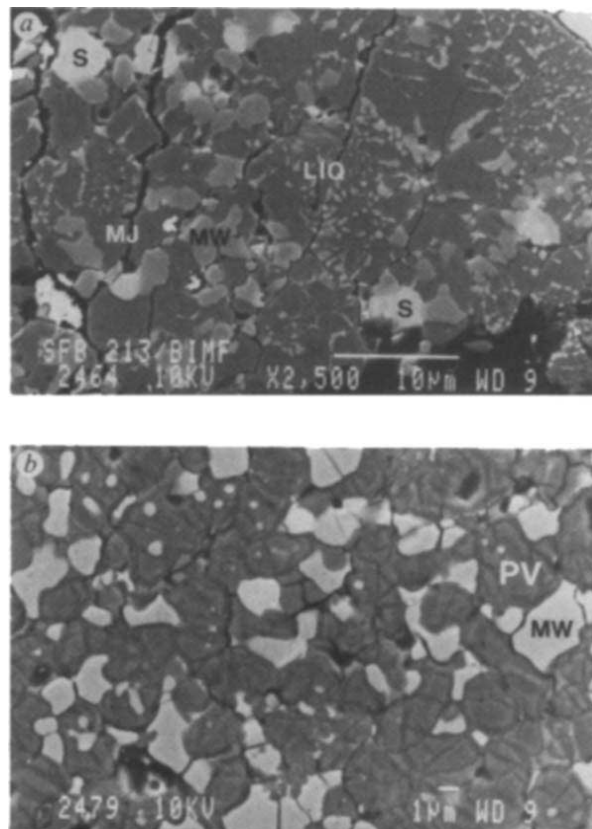


FIG. 1 Scanning electron micrographs (JEOL/JSM 840 A) of the Allende melting experiment (ALHP-3) run at 26 GPa and  $2,050 \pm 50^\circ\text{C}$  for 2 min. Crystalline phases are labelled MW (magnesiowüstite), MJ (majorite garnet) and PV (Mg-perovskite). Quench silicate liquid is labelled LIQ, (Fe,Ni,Re)-sulphide quench liquid is labelled S. *a*, Region where magnesiowüstite (light phase), majorite (dark phase), quench silicate melt (finely mottled patches) and (Fe,Ni)-sulphide (brightest phase) coexist. (Scale bar  $10\ \mu\text{m}$ ). *b*, Cooler region adjacent to *a* with magnesiowüstite (light phase) and Mg-perovskite (dark phase with cleavage and twinning). (Scale bar  $1\ \mu\text{m}$ ). Rhenium heaters and capsules enclosed in 7-mm MgO (5%  $\text{Cr}_2\text{O}_3$ ) octahedra were used because rhenium has a high melting point and is relatively inert in the presence of silicate melt. Low levels of Re contamination in the experimental silicate melts ( $\sim 0.5$  wt%, see Table 1) were detected. The Re capsule interiors had  $\leq 1$  wt% Fe contamination originating from the Allende sample. Microprobe analyses (20-s counting times) indicate that the oxygen concentration in the capsule wall was  $< 0.1$  mol%. Truncation edge length on Toshiba grade 'F' tungsten carbide cubes was 3 mm. High-temperature and pressure calibrations were done by bracketing the  $\text{Mg}_2\text{SiO}_4$  ilmenite/perovskite<sup>28,29</sup> and the  $\text{Mg}_2\text{SiO}_4$   $\beta$ -phase/ $\gamma$ -spinel<sup>30</sup> phase transitions. Starting material for experiments was Allende CV3 provided by M. Prinz of the American Museum of Natural History. Samples were prepared by grinding the meteorite to a powder in an alumina crucible. Temperature was monitored by W3Re/W25Re or Pt S-type thermocouples. In some experiments temperature was estimated from power-temperature calibration curves. All experiments were brought to run pressure at room temperature and then heated. Quenching was accomplished by shutting off the electrical power. Run products were cast in epoxy, sectioned and polished for microscope, microprobe and X-ray diffraction analyses.

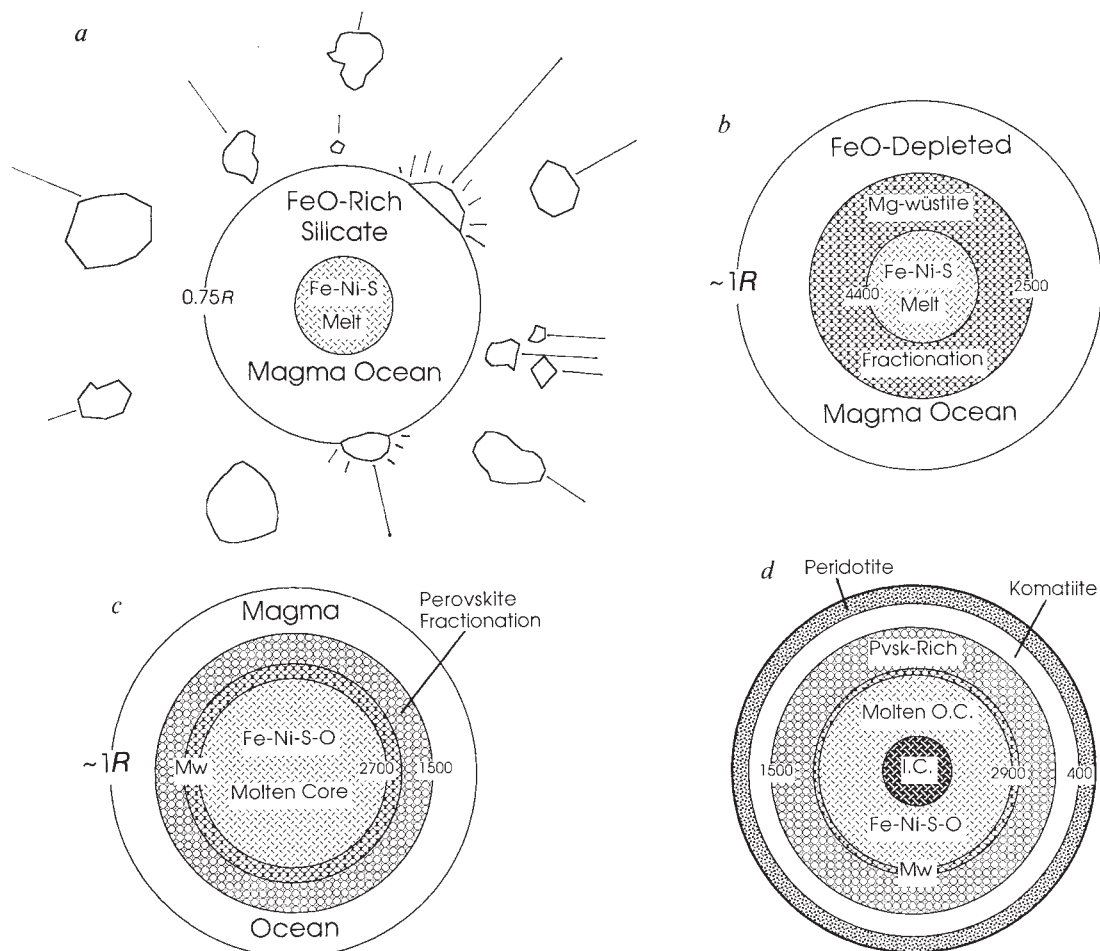
\* Present address: Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA.

The experimental high-pressure phases are constituents that may separate by density differences deep in an extensively molten chondritic Earth. As the chondritic silicate solidus temperature is exceeded during accretionary growth, dense low-temperature Fe-Ni-S melts will first segregate downward to form an early core (Fig. 2a). Later, as the Earth approaches its present radius and cools from temperatures above the silicate liquidus, FeO-rich magnesiowüstite will crystallize. (This conclusion is supported by experimental work<sup>6</sup> on olivine composition at 50 GPa.) Because it should be denser<sup>7</sup> than coexisting silicate melts at all pressures within the Earth, magnesiowüstite will accumulate at deep levels, mantling the Fe-Ni-S proto-core (Fig. 2b). The fate of the FeO-rich magnesiowüstite layer which would mantle the molten Fe-Ni-S core cannot yet be predicted with certainty because of the lack of relevant thermodynamic data at core pressures. Experiments<sup>8</sup> on the system Fe-Ni-S-O at 5 and 16 GPa, however, show a small but increasing solubility of FeO in (Fe, Ni)-sulphide melt with increasing pressure. There is experimental evidence for the metallization of FeO at very high pressure<sup>9,10</sup> and for the increased solubility of FeO in molten Fe with pressure<sup>11-13</sup>. These investigations indicate that incorporation of significant FeO from magnesiowüstite into the (Fe, Ni)-sulphide liquid is possible at outer core pressures. If

so, then magnesiowüstite fractionation may be an important factor in providing a concentrated local FeO source to feed the growing core. Mg-perovskite is also expected to crystallize near the liquidus in the lower mantle<sup>14,15</sup>, but if Mg-perovskite is less dense<sup>16</sup> than melt below 1,680 km ( $P > 70$  GPa), as silicate liquid compressibility studies<sup>17-21</sup> predict, it will not reside as an early basal cumulate. Instead, Mg-perovskite will dissolve during its net movement upward along an adiabat of the convecting magma ocean. Mg-perovskite will become the dominant fractionating phase later, when the lower mantle cools further and solidifies from the bottom up (Fig. 2c).

Mass balance is also consistent with early magnesiowüstite fractionation. If the composition of initial bulk Earth is assumed to be equivalent to that of Allende, then removal of 30% of the magnesiowüstite will decrease the bulk FeO/FeO+MgO from 0.37 to 0.10, equal to that of average upper mantle peridotite (Fig. 3). The addition of this magnesiowüstite to a 7 wt% Fe-Ni-S proto-core (40% of present core volume) produces a core with radius of ~3,884 km. This core volume is ~28% larger than the Earth's present core. If MgO from magnesiowüstite is assumed to be an expelled residue residing in the lower mantle and not in the core, then the core volume is reduced, with a radius approaching 3,484 km. Expelled MgO, if mixed into the whole

FIG. 2 Highly idealized cross-section cartoons at four stages of solid-liquid chondritic Earth differentiation, consistent with experimentally determined phase equilibria and density relations. For simplicity, the labelling of layers indicates only the dominant constituent or state. Crystal fractionation by gravitational settling or flotation was assumed possible and efficient at the sub-liquidus boundary layers of the convecting magma ocean. Effects of degassing or atmosphere production have not been addressed. The timescale from a to d is progressive but not necessarily linear. a, Simplified rendition of the Fe-Ni-S protocore segregation. The abundances in the Allende meteorite was used to estimate the volume proportions of the silicate and Fe-Ni-S melts. Infalling chondritic bodies were explicitly assumed to increase the size of the Earth and implicitly assumed to enhance silicate melting. Although not shown, a Mars-sized object that may have collided with Earth and formed the



Moon<sup>31</sup> could be another source of chondritic material and heat for the proto-Earth. b, Amount of the Earth occupied by Fe-Ni-S proto-core +30% magnesiowüstite fractionation. If all fractionated magnesiowüstite were incorporated in the core, the radius would be ~400 km larger than the present core. The molten silicate mantle is FeO/FeO+MgO=0.10. Magnesiowüstite fractionation and FeO solution in the core, although shown as separate events for clarity, are expected to occur simultaneously. c, Later stage of differentiation, in which most of the FeO from magnesiowüstite is dissolved in the molten core. The dominant fractionating crystalline phase at this time is Mg-perovskite, which enriches the mantle below 700 km in

SiO<sub>2</sub>. Note that the depth of perovskite accumulation occurs in a layer that contains its level of neutral buoyancy (~1,680 km; ref. 21) with silicate melt. d, Mostly solidified mantle after magmatic differentiation. Komatiite layer is a bulk chemical required by mass balance<sup>26</sup> rather than a strictly petrological label. Komatiitic melts may reside here as well as the crystalline phases (depending on the pressure and temperature) olivine, clinopyroxene, majorite,  $\beta$ -phase, and Mg- and Ca-perovskite. With the exception of peridotite, the composition of the layers may not be the same as that in present Earth owing to solid-state convective mixing.



mantle, may lower the bulk Si/Mg and give some concave-downward curvature to the vector in Fig. 3. Mass balance requires the chemical composition of the core to be 75 wt% (Fe, Ni), 13 wt% S and 12 wt% O. The abundances of S and O are somewhat higher than, for example, the 10 wt% oxygen proposed for the core<sup>9</sup> based on its seismically observed density<sup>22</sup> and are most likely overestimates. This is because the present mass balance assumes efficient element incorporation into the core and the resulting concentrations for sulphur and oxygen are therefore the maximum possible values. If, for example, some sulphur is selectively volatilized during accretion or <100% of the FeO in magnesiowüstite dissolves in the core then the light-element concentration may be lower. The present model also assumes, for simplicity, that Earth was formed only of carbonaceous chondrite. Because carbonaceous chondrites are the most oxidized of the chondrites, this model also results in the maximum oxygen in the core. If the model is modified to include some ordinary and enstatite chondrites, then the amount of oxygen in the core is lowered. Although the Allende meteorite may not uniquely represent the composition of bulk Earth, the Allende phase-equilibria results at very high pressure do demonstrate the importance of magnesiowüstite for chondritic Earth differentiation.

Trace-element abundances in the upper mantle are not strictly chondritic<sup>23–25</sup>—for example, the concentrations of Cr, V and Mn are too low. The data for preliminary element partitioning between magnesiowüstite and Allende silicate liquid in Table 1 show that Cr and Mn are compatible elements during melting in the lower mantle. Magnesiowüstite fractionation will have the net effect of depleting the upper portion of the mantle in Cr and Mn and this depletion may be preserved in modern peridotites. Ni also behaves compatibly, but the partition coefficient for magnesiowüstite/Allende silicate melt is several orders of magnitude lower than possible metal-silicate values<sup>24</sup>. Segregation of primordial Ni to the core by magnesiowüstite fractionation would deplete the mantle Ni<sup>+2</sup> concentration, but only to about the levels recorded in peridotites. In contrast, if equilibrium metal-silicate exchange reactions had dominated core formation processes in the early Earth, then complete Ni<sup>+2</sup> extraction from upper mantle olivines and pyroxenes would

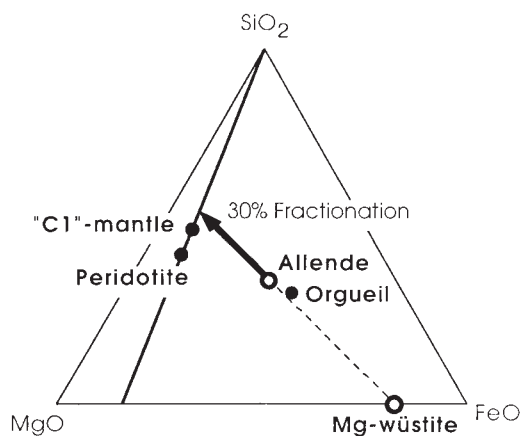


FIG. 3 Ternary diagram with FeO–MgO–SiO<sub>2</sub> in wt% after O'Hara<sup>32</sup>. Solid arrow shows the compositional change of initial Allende silicate melt owing to 30% magnesiowüstite fractionation. Variation of Fe/Mg in magnesiowüstite during fractionation is not yet known and fractionation vector may actually be slightly concave upward. The bold line extending from the SiO<sub>2</sub> apex represents compositions with constant FeO/FeO + MgO = 0.10 such as upper mantle peridotites or 'chondritic mantle'. Magnesiowüstite fractionation of 30% from composition Allende yields a melt with FeO/FeO + MgO = 0.10. Later-stage evolution of the magma ocean involving Mg-perovskite fractionation and olivine addition, producing peridotite from a higher Si/Mg chondritic melt, is described in ref. 26. Position of the Orgueil meteorite (C1) is shown on the compositional diagram for reference.

TABLE 1 Electron microprobe analyses

|                                | Melt (10) | Mg-wüstite (6) | Majorite (4) | Perovskite (5) |
|--------------------------------|-----------|----------------|--------------|----------------|
| SiO <sub>2</sub>               | 38.30     | 0.25           | 50.93        | 49.34          |
| TiO <sub>2</sub>               | 0.16      | 0.06           | 0.06         | 0.21           |
| Al <sub>2</sub> O <sub>3</sub> | 2.14      | 0.19           | 4.42         | 2.52           |
| Cr <sub>2</sub> O <sub>3</sub> | 0.51      | 0.67           | 0.64         | 0.76           |
| FeO                            | 27.49     | 73.36          | 11.60        | 15.75          |
| MgO                            | 25.70     | 19.08          | 26.11        | 27.71          |
| MnO                            | 0.20      | 0.35           | —            | —              |
| CaO                            | 3.24      | 0.19           | 4.07         | 1.07           |
| NiO                            | 0.38      | 1.48           | —            | —              |
| Na <sub>2</sub> O              | 0.44      | 1.37           | 0.20         | 0.25           |
| P <sub>2</sub> O <sub>5</sub>  | 0.18      | 0.00           | 0.28         | 0.00           |
| ReO                            | 0.54      | 0.50           | 0.35         | 0.45           |
| Total                          | 99.28     | 97.50          | 98.65        | 98.06          |

| D (Mg-wüstite/melt)*           |     |
|--------------------------------|-----|
| Cr <sub>2</sub> O <sub>3</sub> | 1.3 |
| FeO                            | 2.7 |
| MnO                            | 1.8 |
| NiO                            | 3.9 |
| Na <sub>2</sub> O              | 3.1 |

Analyses of the products of the 2-min melting experiment at 26 GPa and 2,050 ± 50 °C. Numbers in parentheses indicate the number of analyses. All concentrations are in weight per cent (wt %). Camebax SX 50 electron microprobe; accelerating voltage, 15 kV; sample current, 30 nA.

\* Partition coefficients,  $D = X_{\text{Mg-wüstite}}/X_{\text{melt}}$  where  $X$  is the weight per cent.

have occurred. How these partition coefficients change as core pressures are approached is a subject for future investigation. The partition coefficient for Na between magnesiowüstite and the Allende silicate melt (Table 1) suggests that Na depletion in the upper mantle may not only be due to volatility during accretion.

An idealized sketch of an Earth cross-section describing the redistribution of material after the initial stage of differentiation, based on Allende melting experiments, is shown in Fig. 2d. After core formation and magnesiowüstite fractionation, the FeO-depleted silicate mantle will cool and solidify further, first fractionating Mg-perovskite at depths below 700 km, and then concentrating buoyant olivine in the peridotite shallow mantle. Perovskite subtraction and olivine addition will produce an upper mantle with Si/Mg equivalent to that of average peridotite. Details of this later stage of Earth differentiation are given in ref. 26.

Several questions remain. Is the D' layer at the core/mantle boundary related to magnesiowüstite fractionation? Could it be residual magnesiowüstite-rich material not (yet?) incorporated into the core? Is the inner core devoid of light elements, or could there be an FeO component (see ref. 27) as the present model suggests?

If magnesiowüstite fractionation is important in building planetary cores at very high pressure, then the composition of the cores of smaller bodies with low internal pressures, such as Mercury, may be significantly different than Earth's. Similarly, iron meteorites may be poor analogues for the cores of the larger terrestrial planets Venus and Earth. Mars has high enough internal pressures to stabilize magnesiowüstite, but it is unlikely that significant amounts of FeO from magnesiowüstite can dissolve into Fe–Ni–S melt at its core depths ( $P = 20\text{--}45$  GPa; ref. 23). Mars may therefore have a two-phase core of crystalline magnesiowüstite and molten Fe–Ni–S. □

Received 11 May; accepted 6 July 1990.

- Anders, E. & Grevesse, N. *Geochim. cosmochim. Acta* **53**, 197–124 (1989).
- Jarosewich, E., Clarke, R. S. Jr & Barrows, J. N. (eds) *Smithsonian Cent. Earth Sci.* **27** (1987).
- Jagoutz, E. *et al. Proc. lunar planet. Sci. Conf.* **10**, 2031–2050 (1979).
- Anderson, D. L. & Bass, J. D. *Nature* **320**, 321–328 (1986).
- Ohtani, E., Kato, T. & Sawamoto, H. *Nature* **322**, 352–353 (1986).

6. Williams, Q. *Geophys. Res. Lett.* **17**, 635–638 (1990).
7. Jeanloz, R. & Thompson, A. B. *Rev. Geophys. Space Phys.* **21**, 51–74 (1983).
8. Urakawa, S., Kato, M. & Kumazawa, M. in *High-Pressure Research in Mineral Physics* (eds Manghnani, M. H. & Syono, Y.) 95–111 (Terra Scientific, Tokyo, 1987).
9. Jeanloz, R. & Ahrens, T. J. *Geophys. J. R. astr. Soc.* **62**, 505–528 (1980).
10. Knittle, E. & Jeanloz, R. *Geophys. Res. Lett.* **8**, 1541–1544 (1986).
11. Ohtani, E., Ringwood, A. E. & Hibberson, W. *Earth planet. Sci. Lett.* **71**, 94–103 (1984).
12. Kato, T. & Ringwood, A. E. *Phys. Chem. Miner.* **16**, 524–538 (1989).
13. Knittle, E. & Jeanloz, R. *Geophys. Res. Lett.* **16**, 609–612 (1989).
14. Uto, E. & Takahashi, E. *Nature* **328**, 514–517 (1987).
15. Ohtani, E. & Sawamoto, H. *Geophys. Res. Lett.* **14**, 733–736 (1987).
16. Jeanloz, R. & Knittle, E. *Phil. Trans. R. Soc. A* **328**, 377–389 (1989).
17. Agee, C. B. & Walker, D. *J. geophys. Res.* **93**, 3437–3449 (1988).
18. Agee, C. B. *Eos* **70**, 1418 (abstr.) (1989).
19. Rigden, S. M., Ahrens, T. J. & Stolper, E. M. *Science* **226**, 1071–1074 (1984).
20. Rigden, S. M., Ahrens, T. J. & Stolper, E. M. *J. geophys. Res.* **94**, 9508–9522 (1989).
21. Miller, G. H., Stolper, E. M. & Ahrens, T. J. *J. geophys. Res.* (in the press).
22. Birch, F. *J. geophys. Res.* **69**, 4377–4388 (1964).
23. Ringwood, A. E. *Origin of the Earth and the Moon* (Springer-Verlag, New York, 1979).
24. Wänke, H., Dreibus, G. & Jagoutz, E. in *Archean Geochemistry* (eds Kröner, A., Hanson, G. N. & Goodwin, A. W.) 1–24 (Springer-Verlag, Berlin, 1984).
25. O'Neill, H. St C. *Geochim. cosmochim. Acta* (in the press).
26. Agee, C. B. & Walker, D. *Earth planet. Sci. Lett.* **90**, 144–156 (1988).
27. Jephcoat, A. & Olson, P. *Nature* **325**, 332–335 (1987).
28. Sawamoto, H. in *High-Pressure Research in Mineral Physics* (eds Manghnani, M. H. & Syono, Y.) 209–219 (Terra Scientific, Tokyo, 1987).
29. Ito, E. & Takahashi, E. *J. geophys. Res.* **94**, 10637–10646 (1989).
30. Katsura, T. & Ito, E. Institute for Study of the Earth's Interior, Tech. Rep. **16**, Ser. A (1988).
31. Benz, W., Cameron, A. G. W. & Melosh, H. J. *Icarus* **81**, (1989).
32. O'Hara, M. J., Saunders, M. J. & Mercy, E. L. P. *Phys. Chem. Earth* **9**, 571–604 (1975).

ACKNOWLEDGEMENTS. I thank the staff of the Bayerisches Geoinstitut for technical assistance, W. Reichstein for help with SEM imaging, and R. Jeanloz and H. Palme for reviews.

## Determination of total strain from faulting using slip measurements

C. H. Scholz & Patience A. Cowie

Lamont-Doherty Geological Observatory and Department of Geological Sciences, Columbia University, Palisades, New York 10964, USA

IN the past twenty years it has become commonplace in seismology to sum seismic moments for large earthquakes to determine the contemporary fault slip rate<sup>1</sup> or regional strain rate produced by earthquakes<sup>2</sup>. The method has several drawbacks: it often greatly underestimates deformation rates predicted from plate tectonics, either because the seismic history is of insufficient length or a substantial amount of fault slip is aseismic. Also, it can be used to calculate only current deformation rates, and cannot be applied to earlier geological eras or to estimate total strain. These problems can be overcome by applying the same methods to geological measures of fault displacement. Complete fault data sets, however, are generally not available. Here we show that faults obey general scaling laws in their size frequency distribution and in the relation between displacement and fault length. Combining these scaling relations, we demonstrate that the calculation of strain can be successfully applied to sparse geological data sets, because most strain is produced by the largest faults so that the data set need not be complete for small fault sizes.

The calculation of deformation rates from the sum of seismic moments of earthquakes often yields good results when applied to plate boundaries<sup>1,3–5</sup>. There, slip rates are high enough that the short seismic record is still sufficiently long to characterize the long-term rate. The shortcomings of the method become apparent when it is applied to diffuse zones of deformation, where it often predicts deformation rates that are a small fraction of those predicted from plate tectonics<sup>6,7</sup> and even the directions of maximum shortening may differ significantly from those indicated by plate tectonics<sup>7</sup>. In such regions the slip rates of individual faults may be slow enough that their earthquake recurrence times are much longer than the seismic record, which is therefore a poor sample of the long-term rates and even of the direction of deformation. Furthermore, in many regions a significant portion of fault slip may be aseismic and not represented

in the earthquake record. In principle one could calculate regional strain or strain rate with fault data if one had a sufficiently complete data set of fault length and slip. This would alleviate the problem of the short seismic record and would also include any contribution from aseismic fault slip. This method has been tested for intraplate Japan, where there is both a very complete fault data set and a long seismic history. There, strain rates for the late Quaternary, obtained from the fault data, were found to be consistent with that obtained from the 400-year seismic history<sup>8</sup>. As such a complete data set is not usually available, it would be useful if a procedure applicable to sparse data could be established. This method could then be extended to estimate strain in regions that are no longer seismically active.

Summing seismic moments is practical because the size distribution of earthquakes obeys the relation

$$N(M_0) = aM_0^{-B} \quad (1)$$

where  $N(M_0)$  is the number of earthquakes of seismic moment  $\geq M_0$ , and the moment is the product of shear modulus, slip and fault area,  $\mu uA$ . Because  $B$  has a universal value of  $\sim 2/3$ , the integral of equation (1) is strongly convergent and one need not have a complete catalogue of small earthquakes to obtain a good estimate of total moment release. That equation (1) is convergent is fundamental; seismic moment is proportional to strain and to seismic energy release, both of which must be finite. Because the maximum-size earthquake in a region can always be defined, but the minimum earthquake size is undefined, the convergence of equation (1) must not depend on the minimum-size earthquake detected, that is,  $B$  must be less than 1.

It has been observed that fault populations obey power-law size distributions similar to equation (1). In a given population, the number of faults of length  $\geq L$  is<sup>9–12</sup>

$$N(L) = a_1 L^{-C} \quad (2)$$

and the distribution of accumulated fault displacement  $\geq D$  is<sup>9,13</sup>

$$N(D) = a_1 D^{-C_1} \quad (3)$$

Examples of these distributions are shown in Fig. 1.

Fault displacement and the length of the fault are related. In Fig. 2,  $D$  and  $L$  are plotted for a number of data sets for faults in different tectonic environments. For each individual data set there is a linear relation between  $D$  and  $L$

$$D = \gamma^* L \quad (4)$$

Elsewhere<sup>14</sup> we show that this relation is predicted from a fault growth model in which  $\gamma^*$  is a critical shear strain that is related to the inelastic yield strength of the material in which the fault forms. This relation holds only for faults of finite length and

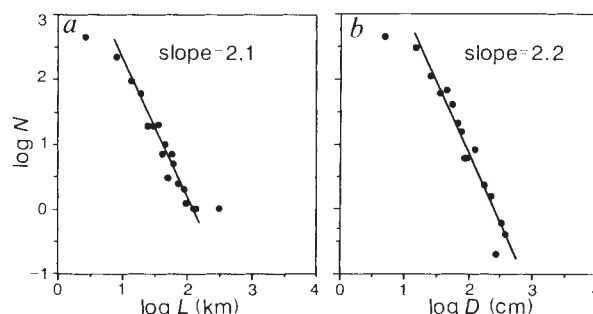


FIG. 1 Fault size distributions. *a*, Number of faults as a function of length for intraplate faults in Japan. Data are from a tabulation in ref. 20, of the original source in ref. 21. *b*, Number of faults as a function of accumulated displacement. The data set is of Neogene faults in the Boso and Miura Peninsulas, Japan<sup>13</sup>. These frequency distributions are the derivatives of the cumulative distributions discussed in the text, so the slopes of the lines are  $-(C+1)$ .



not for plate boundaries. Equation (4) is similar to the slip-length relationship for large earthquakes<sup>15</sup> and is consistent with the idea that fault displacement is built up by the cumulative slip of many earthquakes<sup>14</sup>. Watterson<sup>16</sup> attempted to fit such data with a universal scaling law,  $D \propto L^2$ , but this is inconsistent with both individual data sets and with the collective one. Equation (4) indicates that equations (2) and (3) are therefore equivalent and that  $C = C_1$ , which agrees within the uncertainties of the observed values shown in Fig. 1. Values of  $C$  in the cited work are in the range 1.1–1.6, although values of  $C_1$  as low as 0.4 have been reported<sup>13</sup>.

To determine the total contribution of all faults in an area we integrate over the distributions. Thus the total fault length is obtained by integrating equation (2)

$$\int -\frac{dN(L)}{dL} L dL = \frac{a_1 C}{1-C} L^{1-C} \quad (5)$$

A similar expression is obtained for total fault displacement. To sum fault moment, we have two possibilities. For 'large' faults, which extend through the brittle crust of thickness  $W$ , each fault will have the same width  $W$  (neglecting corrections for dip). In this case their geometric moment (product of displacement and area) will be  $M_g = DLW$ . For 'small' faults, with dimensions  $< W$ , the moment will be  $M_g = DL^2$ . Using equations (2) and (4), the moment sum for the case of large faults ( $L \geq W$ ) is

$$\sum_{k=1}^n M_g^k = a_1 W \gamma^* C \int -\frac{dN(L)}{dL} L^2 dL = \frac{a_1 W \gamma^* C}{2-C} L^{2-C} \quad (6)$$

and for small faults ( $L < W$ )

$$\sum_{k=1}^n M_g^k = a_1 \gamma^* C \int -\frac{dN(L)}{dL} L^3 dL = \frac{a_1 \gamma^* C}{3-C} L^{3-C} \quad (7)$$

In the data sets shown in Fig. 2 the parameter  $\gamma^*$  varies between  $\sim 5 \times 10^{-3}$  and  $5 \times 10^{-2}$ . Because this parameter is related to the

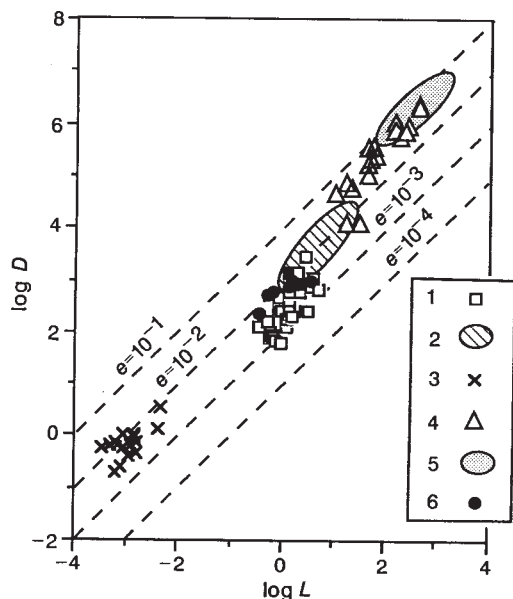


FIG. 2 Logarithmic plot of mean fault displacement (cm) against fault length (km) for data sets from a variety of tectonic environments. Dashed lines are lines of constant shear strain,  $D/L$ . 1, normal faults, British coal measures (ref. 22); 2, normal faults, Lorraine coal fields, France (ref. 9); 3, normal faults, lacustrine sediments, Japan (ref. 23); 4, thrust faults, Scotland (ref. 24); 5, continental strike-slip faults (ref. 25); 6, normal faults, Iceland rifts (ref. 26). Mean fault displacement is taken to be 2/3 of the maximum displacement, which is the usual parameter tabulated in the original sources. This is based on an elastic-fault model. If an inelastic model is assumed, such as in ref. 14, mean displacement would be  $\sim 1/2$  of the maximum displacement.

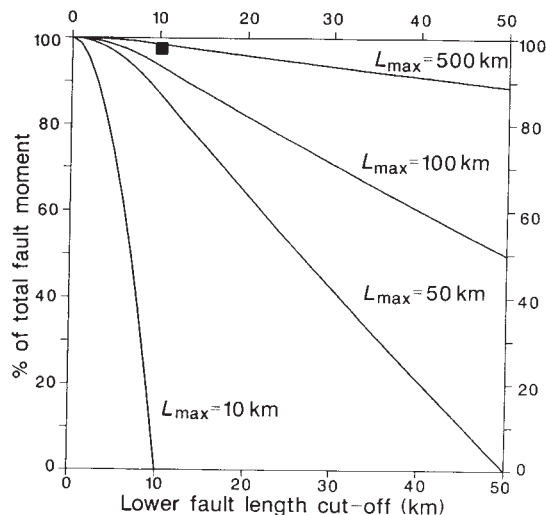


FIG. 3 The effect of a limited fault data set on moment sums. The percentage of total fault moment is plotted as a function of the lower fault length cut-off in a data set for the value of  $C$  shown in Fig. 1, 1.1, and for four assumed values of  $L_{\max}$ .  $W$  is assumed to be 10 km, so equation (6) is used above 10 km and equation (7) is used below 10 km. The solid square is for the Japanese fault data set used in Fig. 1.

yield strength of the material<sup>14</sup>, it varies with rock type and tectonic environment and has to be determined for the region of interest. Once the displacement-length relation is calibrated, the total contribution from all faults in the region can be determined just from their length.

Because the moment must be finite, as discussed above, equations (6) and (7) must converge, that is,  $C < 2$  and  $C < 3$ , respectively. The typical values for  $C$  in the cited work are between 1.1 and 1.6 so that total fault moment, and thus total fault strain, are indeed finite and only sensitive to the largest faults in a region. This contrasts with the idea<sup>17</sup> that significant amounts of strain may be hidden in small faults below the observational limit. This is illustrated in Fig. 3 where, for the value of  $C$  in Fig. 1 and several assumed values of maximum fault length  $L_{\max}$ , we show the percentage of total moment that would be contained in a data set complete to various lower fault length cut-offs. Thus if  $L_{\max}$  is 100 km, a data set complete for faults of length  $> 20$  km would contain 80% of the moment, and if  $C$  is known we can correct for the moment contained in faults below the resolution of the data. But the total fault length or displacement, equation (5), are not convergent unless  $C < 1$ . Thus for  $C > 1$  the sum of fault displacements in a region will only be finite if a lower size limit is specified and will be very sensitive to this lower limit.

To calculate the total strain owing to faulting in a region of area  $A$ , we write the moment as a tensor, using information on fault orientations and displacement vectors, and sum over  $n$  faults using Kostrov's<sup>2</sup> formula

$$\epsilon_{ij} = \frac{1}{2WA} \sum_{k=1}^n M_{gij}^k \quad (8)$$

The analysis of a fault data set using equation (8) is straightforward. For example, by assuming some of the relationships demonstrated here, Marrett and Allmendinger<sup>18,19</sup> discuss graphical and numerical methods for doing so. The convergence of equation (6) means that an estimate of total fault strain can be obtained by considering only the largest faults in a region and allows one to determine how good the estimate is.  $\square$

Received 14 March; accepted 24 May 1990.

1. Brune, J. *J. geophys. Res.* **73**, 777–784 (1968).

2. Kostrov, B. *Izv. Acad. Sci. USSR Phys. Solid Earth* **1**, 23–44 (1974).

3. Davies, G. & Brune, J. *Nature (Phys. Sci.)* **229**, 101–107 (1971).
4. Molnar, P. *Bull. seism. Soc. Am.* **69**, 115–134 (1979).
5. Sykes, L. & Quittmeyer, R. in *Earthquake Prediction: an International Review* M. Ewing Ser. 4. (eds Simpson, D. & Richards, P.) 217–247 (American Geophysical Union, Washington, DC, 1981).
6. Jackson, J. & McKenzie, D. *Geophys. J. R. astr. Soc.* **93**, 45–73 (1988).
7. Ekstrom, G. & England, P. *J. geophys. Res.* **94**, 10231–10257 (1989).
8. Wesnousky, S., Scholz, C. & Shimazaki, K. *J. geophys. Res.* **87**, 6829–6852 (1982).
9. Villemin, T. & Sunwoo, C. *C. r. hebdo. Acad. Séanc. Sci. Paris, Ser. 2* **305**, 1309–1311 (1987).
10. Hirata, T. *Pure appl. Geophys.* **131**, 157–170 (1989).
11. Gudmundsson, A. *Tectonophysics* **139**, 295–308 (1987).
12. Gudmundsson, A. *J. struct. Geol.* **9**, 61–69 (1987).
13. Kakimi, T. *Bull. geol. Soc. Jap.* **31**, 467–487 (1980).
14. Cowie, P. & Scholz, C. *Eos* **71**, 631 (1990).
15. Scholz, C. *Bull. seism. Soc. Am.* **72**, 1–14 (1982).
16. Watterson, J. *Pure appl. Geophys.* **124**, 366–373 (1986).
17. Kautz, S. & Sclater, J. *Tectonics* **7**, 823–832 (1988).
18. Marrett, R. & Allmendinger, R. *Geol. Soc. Am., Abstr. Prog.* **20**, A319 (1988).
19. Marrett, R. & Allmendinger, R. *J. struct. Geol.* (in the press).
20. Wesnousky, S. G. thesis, Columbia Univ. (1982).
21. Working Group for Active Faults in Japan *Active Faults in Japan: Sheet maps and inventories* (University of Tokyo Press, 1980).
22. Walsh, J. & Watterson, J. *J. struct. Geol.* **9**, 1039–1046 (1987).
23. Muraoka, H. & Kamata, H. *J. struct. Geol.* **5**, 483–495 (1983).
24. Elliott, D. *Phil. Trans. R. Soc. A* **283**, 289–312 (1976).
25. Ranalli, G. *Tectonophysics* **37**, 1–7 (1977).
26. Ophelm, J. & Gudmundsson, A. *Geol. Soc. Am. Bull.* **101**, 1608–1622 (1989).

ACKNOWLEDGEMENTS. We thank C. Barton, M. Anders and D. Simpson for comments. This work was supported by the NSF. This is Lamont-Doherty Geological Observatory Contribution 4644.

## Fertilization of the desert soil by rock-eating snails

Clive G. Jones\* & Moshe Shachak†\*

\*Institute of Ecosystem Studies, The New York Botanical Garden, Mary Flagler Cary Arboretum, Millbrook, New York 12545, USA

†Mitrani Center for Desert Ecology, Jacob Blaustein Institute for Desert Research, Ben-Gurion University of the Negev, Sede Boqer Campus, Israel 84990

**PLANT productivity in deserts can be nitrogen-limited. Nitrogen inputs are often low, soil nitrogen pools small, and losses from runoff, erosion, volatilization and denitrification, can be high<sup>1–6</sup>. We have now found an unusual, important source of soil nitrogen in the Central Negev Highlands of Israel, a limestone rock desert with patches of soil (Fig. 1). Snails feed on endolithic lichens that grow within the rock, ingesting both rock and lichens<sup>7,8</sup>, and depositing their faeces on the soil under the rocks. Snails transfer between 22–27 mg nitrogen per m<sup>2</sup> per year to soil, which constitutes about 11% of total soil nitrogen inputs, at least 18% of net soil inputs, and a minimum of 27% of the nitrogen annually accumulated by endolithic lichens from dust. The substantial contribution of snails to the nitrogen cycle is probably important for higher plant production.**

FIG. 1 The rocky Central Negev Desert Highlands, Israel, showing higher plants growing in patches of soil adjacent to limestone rock. The two substrates support different plant communities. Endolithic lichens live within the upper few mm of the limestone rocks, which cover 70% of the desert area; they are the dominant cryptogamic plants in this desert and have a low standing crop<sup>22,23</sup>. The soil patches, about 30% of the desert area, support a relatively large biomass of higher plants<sup>10,11</sup>. The endolithic lichen-rock layer contains about 0.5% nitrogen, which is taken up by the lichens from dust deposition<sup>9</sup>. Dew forms on rocks, providing the moisture for nitrogen uptake from dust<sup>24</sup>. Snails of the genus *Euchondrus* ingest the endolithic lichens and the rock within which they grow<sup>7,8</sup>. Snail faeces are deposited on the soil under the rocks, where the roots of higher plants grow. Nitrogen input to the soil from snail faeces is a principal component of the nitrogen cycle of this desert. (Photograph, C.G.J.).

Three snail species in the Negev, *Euchondrus albulus* Mousson, *E. desertorum* Roch and *E. ramonensis* Granot, cut feeding trails in rocks<sup>7,8</sup>, converting rock to soil at a rate between 69.5–110.4 g per m<sup>2</sup> per year—equivalent to that due to wind-borne dust deposition<sup>7</sup>. Snails feed on the rocks when dew is present (night and early morning), sheltering under the rocks during the day, and depositing their faeces on the soil before feeding again<sup>7</sup>. We measured faecal production and nitrogen content of naturally occurring densities of the dominant snail species, *E. albulus*<sup>7</sup>, in the laboratory and field, under conditions in which the number of days of natural and/or artificial dew varied (Table 1).

The number of dew days was a significant factor determining faecal production per snail per day in two experiments (Table 1, A1). When data were expressed as faecal production per snail per dew day, there was no significant difference in faecal production between the two experiments (Table 1, B1), or between laboratory and field treatments in experiment 1 (Table 1, B2). Laboratory values were significantly greater than the two field values in experiment 2 (Table 1, B3). However, none of the field treatments in either experiment differed significantly from one another, with faecal production values varying by about 12% (Table 1, B4), a remarkably consistent faecal production rate.

Faecal nitrogen ranged between 0.40–0.52% dry weight (DW) (Table 1), and was not significantly different between laboratory and field treatments (Table 1, C1), was not related to the number of dew days (Table 1, C2), but was significantly greater, by 0.1% DW, in experiment 2, compared with experiment 1 (Table 1, C3). Faecal nitrogen was 0.06% DW lower than that of the endolithic lichen-rock layer, indicating that some portion of the ingested nitrogen was being retained for snail growth and maintenance (Table 1, D1). Nitrogen content of the rock interior was significantly lower than that of either the rock-lichen surface layer or snail faeces (Table 1, E1, E2). Fossil nitrogen in the upper surface of the sedimentary limestone contributed less than 10% of the nitrogen of the rock-endolithic lichen layer.

Snails therefore removed nitrogen from the rock-endolithic lichens, excreting a substantial portion of this nitrogen in their faeces. Field estimates of soil nitrogen input due to *E. albulus* ranged from 21.9–26.9 (average 24.3) mg nitrogen per m<sup>2</sup> per year (Table 1). This estimate does not include inputs from other species of rock-eating snails, one of which (*E. desertorum*) is locally abundant<sup>7</sup>.

Estimates of the relative importance of snail-induced nitrogen cycling were made by comparing these data with the soil budget of nitrogen in the same watersheds<sup>5,6,9</sup> (Fig. 2). Snail faeces are estimated to represent 11% of total soil nitrogen input, and at least 18% of net soil nitrogen input. The latter estimate is





TABLE 1 Transfer of nitrogen to soil by snails

| Experiment | Treatment | Lab. or field | No. of dew days | Type of dew            | Faecal production per day (mg DW faeces snail <sup>-1</sup> day <sup>-1</sup> ) (s.e.m.) | Faecal production per dew day (mg DW faeces snail <sup>-1</sup> dew day <sup>-1</sup> ) (s.e.m.) | Faecal nitrogen (% DW) (s.e.m. = analytical error) | Estimated soil nitrogen input (mg total nitrogen m <sup>-2</sup> year <sup>-1</sup> ) |
|------------|-----------|---------------|-----------------|------------------------|------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|----------------------------------------------------|---------------------------------------------------------------------------------------|
| 1          | 1         | Field         | 8               | Natural and artificial | 1.02 (0.11)                                                                              | 1.28 (0.14)                                                                                      | 0.40 (0.01)                                        | 22.4                                                                                  |
| 1          | 2         | Field         | 4               | Natural                | 0.50 (0.05)                                                                              | 1.26 (0.13)                                                                                      | 0.40 (0.01)                                        | 21.9                                                                                  |
| 1          | 3         | Lab.          | 8               | Artificial             | 1.18 (0.12)                                                                              | 1.48 (0.15)                                                                                      | 0.40 (0.01)                                        | 26.0                                                                                  |
| 2          | 1         | Field         | 8               | Natural and artificial | 0.96 (0.07)                                                                              | 1.20 (0.09)                                                                                      | 0.51 (0.03)                                        | 26.9                                                                                  |
| 2          | 2         | Field         | 1               | Natural                | 0.11 (0.01)                                                                              | 1.13 (0.14)                                                                                      | 0.52 (0.01)                                        | 26.0                                                                                  |
| 2          | 3         | Lab.          | 8               | Artificial             | 1.49 (0.11)                                                                              | 1.86 (0.14)                                                                                      | 0.47 (0.01)                                        | 38.3                                                                                  |
| 2          | 4         | Lab.          | 0               | None                   | 0.08 (0.03)                                                                              | —                                                                                                | 0.49 (0.02)                                        | —                                                                                     |

Faecal production (mean mg DW faeces per snail per day and mean mg DW faeces per snail per dew day,  $\pm$ s.e.m.), faecal nitrogen content (mean % DW total nitrogen  $\pm$ s.e.m. of 3 subsamples from pooled faeces by treatment, s.e.m. is analytical error) and estimated annual nitrogen input to soil (mg total nitrogen per m<sup>2</sup> per year) due to *E. albulus* feeding on endolithic lichens. Ten replicate rocks (mean area  $\pm$ s.e.m. =  $71.4 \pm 3.2$  cm<sup>2</sup>) for each treatment in each experiment were collected in November and December, 1989, at the time of seasonal snail activity (September–June<sup>7</sup>), from Ramat Avdat, Central Negev Highlands. Five snails of similar weight (mean fresh weight + shell  $\pm$ s.e.m. =  $339.8 \pm 5.1$  mg snail<sup>-1</sup>), collected at the same time and place as rocks, were added to each rock. The snail density per rock was that found in the field<sup>7</sup>. Each rock with snails was placed in a 17.5 × 28.5 × 8.5 cm, pre-washed plastic arena with a mesh cover. Two experiments (1, 2) were carried out, each with simultaneous runs over 10 days in the laboratory and the field. Experiments were carried out at ambient conditions (Laboratory: 10–20 °C range, 15 °C mean; 70% relative humidity (RH) mean, 52–98% RH range. Field: 5–27 °C range, 14 °C mean; 72% RH mean, 44–100% RH range). Naturally occurring dewfall in the field was used, recorded at the local weather station, either without (experiment 1, treatment 2–4 dew days; experiment 2, treatment 1–2 dew day) or with supplements of artificial dew (experiment 1, treatment 1; experiment 2, treatment 1—both 8 dew days). Artificial dew supplements consisted of 1.5 ml deionized water sprinkled evenly over the rock surface twice daily at 09:00 and 14:00. The same artificial dew regime was used in the laboratory (experiment 1, treatment 3–8 dew days; experiment 2, treatment 3–8 dew days) during the 10-day experiments, dew being allowed to evaporate under the ambient conditions. In experiment 2, treatment 4, ambient relative humidity only (mean 70%; 0 dew days) was used. Faeces were collected from each replicate arena at the end of the experiments, dried and weighed. Faeces from each treatment were then pooled across replicates, homogenized, and 3 subsamples (4–12 mg) withdrawn and analysed for total N by elemental analysis (Carlo-Erba CNS,  $\pm 0.01\%$  DW) using acetonitrile as a calibration standard. Nitrogen content of the endolithic lichen–rock layer consumed by snails was estimated by removing the entire top 3 mm of 4 rocks of the same size as above, collected in the same area, using an engraving tool to simulate snail feeding, followed by elemental analysis of each replicate rock. Nitrogen content of the rock interior (sedimentary N) was determined by splitting rocks vertically in half and sampling the interior below the endolithic lichens, 1 cm down from the surface, followed by elemental analysis of each replicate rock ( $n=5$  rocks of the same size as above, collected in the same area). Estimates of total nitrogen input to soil were obtained by multiplying mg DW faeces per snail per dew day by the faecal nitrogen content for each treatment, and multiplying the resulting value by field snail density (21 per m<sup>2</sup>) and average number of dew days per year (210) based on previous data<sup>7</sup>. Estimates of the importance of snail faecal nitrogen (Fig. 2) used data from field treatments only.

Statistical comparisons: **A. Faecal production per day:** (1) Main effect, dew days (all experiments and treatments). 1-way ANOVA: d.f. = 3,  $F=59.23$ ,  $P<0.0001$ . Scheffe  $F$  a posteriori comparisons and  $P$  values: 0 vs 1 dew day,  $F=0.021$ ,  $P>0.05$ ; not significant (NS); 0 vs 4 dew days,  $F=3.33$ ,  $P<0.05$ ; 0 vs 8 dew days,  $F=34.84$ ,  $P<0.05$ ; 1 vs 4 dew days,  $F=28.26$ ,  $P<0.05$ ; 1 vs 8 dew days,  $F=32.71$ ,  $P<0.05$ ; 4 vs 8 dew days,  $F=12.91$ ,  $P<0.05$ . **B. Faecal production per dew day:** (1) Main effect, experiment (excluding experiment 2, treatment 4–0 dew days). 1-way ANOVA: d.f. = 1,  $F=0.28$ ,  $P=0.60$ , NS. (2) Main effect, lab. vs field, experiment 1. 1-way ANOVA: d.f. = 2,  $F=0.77$ ,  $P=0.47$ , NS. (3) Main effect, lab. vs field, experiment 2 (excluding lab. treatment 4–0 dew days). 1-way ANOVA: d.f. = 2,  $F=10.30$ ,  $P<0.005$ . treatment 1 (field) vs treatment 3 (lab.): Scheffe  $F=6.93$ ,  $P<0.05$ . treatment 2 (field) vs treatment 3 (lab.): Scheffe  $F=8.44$ ,  $P<0.05$ ; treatment 1 (field) vs treatment 2 (field): Scheffe  $F=0.07$ , NS at  $P=0.05$ . (4) Main effect, field (experiments 1 and 2). 1-way ANOVA: d.f. = 1,  $F=0.13$ ,  $P=0.72$ , NS. **C. Faecal nitrogen:** (1) Main effect, lab. vs field (experiments 1 and 2). 1-way ANOVA: d.f. = 1,  $F=0.02$ ,  $P=0.90$ , NS. (2) Faecal nitrogen % DW vs number of dew days (experiments 1 and 2). Regression: d.f. = 6,  $r^2=0.17$ ,  $F=1.04$ ,  $P=0.36$ , NS. (3) Main effect, experiment. 1-way ANOVA: d.f. = 1,  $F=44.56$ ,  $P=0.0011$ . Mean % DW faecal nitrogen  $\pm$ s.e.m.: experiment 1 =  $0.40 \pm 0.001$ ; experiment 2 =  $0.50 \pm 0.013$ . **D. Rock–endolithic lichen layer nitrogen:** (1) Compared to faecal nitrogen. Mean rock–endolithic layer total nitrogen, % DW  $\pm$ s.e.m. =  $0.51 \pm 0.01$ ,  $n=4$ . Mean faecal nitrogen, % DW  $\pm$ s.e.m., across experiments and treatments =  $0.45 \pm 0.02$ ,  $n=7$ . 1-way ANOVA: d.f. = 1,  $F=3.665$ ,  $P=0.09$ . **E. Rock interior nitrogen:** (1) Compared to rock–endolithic lichen layer nitrogen. Mean rock interior total nitrogen, % DW  $\pm$ s.e.m. =  $0.05 \pm 0.01$ ,  $n=5$ ; versus rock–endolithic lichen layer (see D1): d.f. = 7,  $t$  2-tailed = 28.85,  $P<0.0001$ . (2) Compared to faecal nitrogen (see D1): d.f. = 10,  $t$  2-tailed = 17.25,  $P<0.0001$ .

conservative because losses due to volatilization, denitrification and erosion are unknown, and because net dry deposition is assumed to equal total dry deposition (whereas it is almost certainly less, because some deposition is due to local dust redistribution rather than importation of dust from other regions<sup>9</sup>). The net input estimate takes into account the annual net losses of nitrogen (in rain and dust, and from soil–crust nitrogen fixation) following rain-induced runoff from rocks and soil<sup>5,6</sup>. But snail faeces are retained in the system following runoff events, because they are deposited under rocks and are protected by plugs of soil that accumulate on the upslope side of the rock–soil interface.

We estimate that 27% of the nitrogen that is annually accumulated by endolithic lichens from dust, is transferred to the soil by snails (Fig. 2). Dry deposition of dust on rocks is the main source of nitrogen for endolithic lichens<sup>9</sup> because these lichens do not fix atmospheric nitrogen<sup>5,6</sup>; fossil sedimentary nitrogen is low, and rainfall nitrogen inputs to rock are small because

rainfall is low (<90 mm per year), and infiltration of rain into rock is minimal<sup>5,6</sup>. The estimate of nitrogen transfer is conservative because the calculation is based on total dry deposition (rather than a lesser, but unknown net deposition), and because the calculation assumes that lichens take up all of the dust nitrogen (rather than some unknown portion of that deposited on rocks).

Snail-induced cycling of nitrogen from cryptogamic plants in the rocks to soil may influence productivity of higher plants growing in the desert soil. Snail faeces are deposited under the rocks, where the roots of higher plants grow. The rock–soil interface is a high-quality microhabitat for plant roots, because of the relatively higher soil moisture regime, compared to the uncovered soil<sup>10,11</sup>. Nitrogen may be limiting to higher plants in these microhabitats, because water availability is greatest here. Snail faeces are therefore being deposited in a micro-environment where they can most probably influence higher plant production.

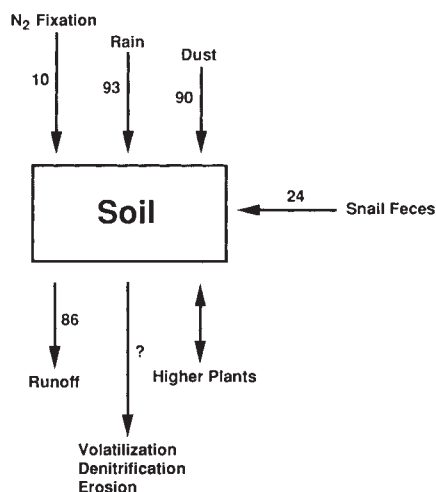


FIG. 2 Inputs and outputs of nitrogen for patches of soil in the Negev Desert Highlands watersheds. Values are average mg total nitrogen per m<sup>2</sup> per year. Data on nitrogen budgets are from studies carried out in 1984–1987<sup>5,6,9</sup>. Rain water (68 and 118 mg nitrogen per m<sup>2</sup> per year, 1984–1985 and 1985–1986, respectively, average 93) for nitrogen determination was collected 1 m above the desert surface throughout 1984–1986. The bulk collector was cleaned of dust every 2 days and washed when clouds indicated a potential rain event. Nitrate was determined with NAS reagent, ammonium with indophenol blue, and nitrites with a modified Gries method<sup>25</sup>. Runoff from rocks and soil is a net annual loss of nitrogen. After each watershed runoff event (20 and 153 mg nitrogen per m<sup>2</sup> per year, 1984–1985 and 1985–1986, respectively, average 86), water samples were taken from a network of pools behind a set of stage recorders, filtered and analysed for NO<sub>3</sub> and NH<sub>4</sub><sup>5,6</sup>. Unfiltered samples were analysed for dissolved and particulate organic nitrogen by the Kjeldahl method. Dust samples (total dry deposition) were collected on a bulk collector, monthly through 1987, weighed and analysed for NO<sub>3</sub>, NH<sub>4</sub> and total nitrogen, as above<sup>9</sup> (monthly range 2.9–10.7, average 7.5 mg total N per m<sup>2</sup>; annual estimate 90 mg N per m<sup>2</sup>). Maximum potential for nitrogen fixation was determined by acetylene reduction<sup>26</sup> for samples of soil crust and rocks along transects in 1986. No detectable nitrogen fixation occurred on rocks<sup>5,6</sup>, and endolithic lichens do not generally fix nitrogen, on the basis of studies in other desert and antarctic environments<sup>27</sup>. The actual amounts of fixation due to soil crust algae-bacteria is probably less than the small amount shown, because the acetylene reduction method tends to overestimate fixation<sup>28,29</sup>. Nitrogen losses due to volatilization, denitrification and erosion are unknown. Estimates of the importance of snail faecal nitrogen inputs to soil were calculated as follows. Percentage total soil nitrogen input (11%)=[snail input (average, 24)/{snail input (24)+rain (93)+soil crust nitrogen fixation (10)+total dry deposition (90)}]×100. Percentage net soil nitrogen input (18%)=[snail input (24)/{snail input (24)+rain (93)+soil crust nitrogen fixation (10)+total dry deposition (90)–runoff (86)}]×100. Percentage snail transfer of rock nitrogen (27%)=[snail input (24)/total dry deposition (90)]×100.

Our study demonstrates the importance of invertebrate consumers in nitrogen cycling, and suggests that ecosystem productivity in this arid environment would be less in the absence of the herbivores. There is considerable debate over the importance of consumers in nutrient cycling<sup>12–18</sup>. This study, together with studies in other deserts on the role of insects such as ants, termites and cicadas<sup>19–21</sup>, show that invertebrates can have a critical role. □

Received 22 May; accepted 11 June 1990.

1. Skujins, J. *Ecol. Bull.* **33**, 477–491 (1981).
2. West, N. & Skujins, J. (eds) *Nitrogen in Desert Ecosystems* (Dowden, Stroudsburg, Pennsylvania, 1978).
3. Boring, L. R., Swank, W. T., Waide, J. B. & Henderson, G. S. *Biogeochemistry* **6**, 119–159 (1988).
4. West, N.E. in *Semi Arid Lands and Deserts: Soil Resource and Rehabilitation*, (ed. Skujins, J.) (Dekker, New York, in the press).
5. Esser, U. thesis, Wilhelms-Universität, Münster (1989).
6. Esser, U., Schreiber, K. F., Shachak, M. & Yair, A. *Trans. 13th Congr. Int. Soc. Soil Science* **3**, 730–731 (1986).
7. Shachak, M., Jones, C. G. & Granot, Y. *Science* **236**, 1098–1099 (1987).
8. Granot, Y. *Israel J. Zool.* **34**, 205–211 (1987).
9. Offer, Z. & Goossens, D. *J. Arid Envir.* **17**, 1–19 (1990).

10. Evanari, M., Shanan, L. & Tadmor, N. *The Negev* 2nd edn (Harvard University Press, Cambridge, 1982).
11. Hillel, D. *Negev. Land, Water and Life in a Desert Environment* (Praeger, New York, 1982).
12. Chew, R. M. *Ohio J. Sci.* **74**, 359–370 (1974).
13. Mattson, W. J. & Addy, N. D. *Science* **190**, 515–522 (1975).
14. Kitchell, J. F. et al. *Bioscience* **29**, 28–34 (1979).
15. Owen, D. F. *Oikos* **35**, 230–235 (1980).
16. Petelle, M. *Oikos* **35**, 127–128 (1980).
17. Choudry, D. *Oikos* **43**, 401–402 (1984).
18. Choudry, D. *Oikos* **45**, 287–290 (1985).
19. Lafage, J. P. & Nutting, W. L. in *Production Ecology of Ants and Termites* (ed. Brian, M. V.) 165–232 (Cambridge University Press, 1977).
20. Crawford, C. S. *J. Arid Envir.* **2**, 31–42 (1979).
21. Schaefer, D. A. & Whitford, W. G. *Oecologia* **48**, 277–283 (1981).
22. Danin, A. & Garty, J. *Z. Geomorphol.* **27**, 423–444 (1983).
23. Kushnir, E. thesis, Univ. Tel Aviv (1979).
24. Kidron, G. thesis, Hebrew Univ. of Jerusalem (1988).
25. *Deutsche Einheits zur Wasseruntersuchung* Vol. 1 (Verlag Chemie, Weinheim, 1974).
26. Hardy, R. W. F., Burns, R. C. & Holsten, R. D. *Soil Biol. Biochem.* **5**, 47–81 (1973).
27. Friedmann, E. I. & Kibler, A. P. *Microb. Ecol.* **6**, 95–108 (1980).
28. Weaver, R. W. in *Field Measurement of Dinitrogen Fixation and Denitrification*, (eds Hauck, R. D. & Weaver, R. W.) 1–9 (Soil Sci. Soc. Amer. Spec. 18, Madison, 1986).
29. Paul, E. A. & Clark, F. E. *Soil Microbiology and Biochemistry* (Academic, San Diego, 1989).

ACKNOWLEDGEMENTS. We thank S. Rozin, S. Nathanson, S. Brand, I. Augie and D. Schmidt for technical assistance, Z. Offer for additional data on dust deposition and dewfall, J. Lawton, G. Lovett, G. Likens, D. Strayer, N. West for critical comment, and the Mary Flagler Cary Charitable Trust for financial support, including visiting scientist support for M. S. at IES. This research was partially supported by a US-Israel Binational Science Foundation grant (M.S. and C.G.). Publication no. 106 from the Mitrani Center. Contribution to the program, Institute of Ecosystem Studies, The New York Botanical Garden.

## Hormonal control of Mg<sup>2+</sup> transport in the heart

A. Romani & A. Scarpa

Department of Physiology and Biophysics,  
Case Western Reserve University, School of Medicine,  
Cleveland, Ohio 44106, USA

**MAGNESIUM is abundant in the mammalian body and the second most abundant cation in cells<sup>1,2</sup>. Because the concentration of intracellular free Mg<sup>2+</sup> is relatively high (0.2–1 mM)<sup>1,3</sup>, Mg<sup>2+</sup> is unlikely to act as a second messenger, like Ca<sup>2+</sup>, by rapidly changing its cytosolic concentration. But changes in Mg<sup>2+</sup> do have profound effects on cellular metabolism, structure and bioenergetics<sup>3–7</sup>. Key enzymes or metabolic pathways<sup>1,6–8</sup>, mitochondrial ion transport<sup>6,9–11</sup>, Ca<sup>2+</sup> channel activities in the plasma membrane and intracellular organelles<sup>3,12,13</sup>, ATP-requiring reactions, and structural properties of cells<sup>4,6,14</sup> and nucleic acids<sup>1,3,7</sup> are modified by changes in Mg<sup>2+</sup> concentration. Yet, although some information is available from giant cells<sup>3,15</sup> and bacteria<sup>3,16,17</sup>, little is known about the regulation of intracellular Mg<sup>2+</sup> in mammalian cells. Here we report a new transport mechanism for Mg<sup>2+</sup> across the sarcolemma of cardiac cells in both intact hearts and dissociated myocytes. We show that noradrenaline, through  $\beta$ -adrenergic stimulation and increase of cyclic AMP, stimulates a large efflux of Mg<sup>2+</sup> from cardiac cells. This transport is of major dimensions and can move up to 20% of total cellular Mg<sup>2+</sup> within a few minutes.**

Figure 1 shows the Mg<sup>2+</sup> efflux from a perfused rat heart. Every 20 s perfusate was collected and assayed for Mg<sup>2+</sup> by atomic absorbance spectrophotometry. Ten minutes before measurements began, the perfusion medium was shifted to one containing zero Mg<sup>2+</sup>. In controls (Fig. 1a), the Mg<sup>2+</sup> content in the perfusate was 7.5  $\mu$ M progressively decreasing to 2  $\mu$ M over 30 min perfusion. By contrast, the addition of noradrenaline (NA) to the perfusate results in a large efflux of Mg<sup>2+</sup> from the heart (Fig. 1b). In different hearts this efflux was stimulated by NA in the perfusion medium in a dose-dependent fashion and was observable over a dose range 0.04–10  $\mu$ M. Qualitatively similar results (not shown) were obtained using adrenaline and isoprenaline.

Figure 2a shows Mg<sup>2+</sup> efflux from perfused hearts after successive short stimulations with 0.2  $\mu$ M NA. Each stimulation



evoked a progressively smaller efflux. This and the previous results could be accounted for by either a down-regulation of noradrenaline receptors or a depletion of magnesium from intracellular pool(s).

As the plasma membrane of cardiac cells contains both  $\alpha$ - and  $\beta$ -adrenergic receptors<sup>18,19</sup>, we used specific  $\alpha$ - and  $\beta$ -antagonists<sup>19-21</sup> to identify which type(s) of receptors were involved in the stimulation of  $Mg^{2+}$  efflux by NA.

Alpha-antagonists such as phentolamine, prazosin and yohimbine had little effect in preventing NA stimulation even at a concentration 10 times greater than the concentration of agonist (not shown). By contrast, propranolol (a nonselective  $\beta$ -antagonist) completely blocks the  $Mg^{2+}$  efflux induced by NA in perfused hearts (Fig. 2b).

A series of controls (not shown) determined that  $Mg^{2+}$  efflux was specific to adrenergic stimulation and was not secondary to an increase in heart rate or/and positive inotropic effects. Pacing the hearts through A-V node stimulation at 100 or 150

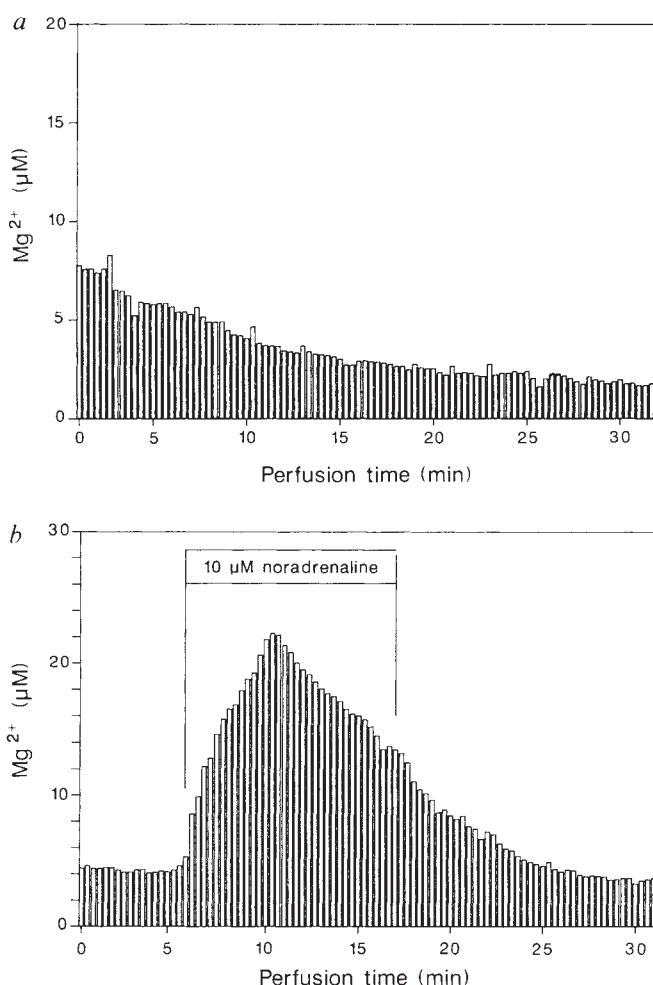


FIG. 1  $Mg^{2+}$  efflux in control perfused hearts (a) and after addition of NA (b). One experiment typical of three, both for control and NA-stimulated hearts is shown.

**METHODS.** Male Sprague-Dawley rats (190–230 g) were used as heart donors after anaesthetic treatment. Hearts (1.3–1.4 g) were removed and perfused with a buffer containing (mM): 120 NaCl, 3 KCl, 10 HEPES buffer, 10 glucose, 1.2  $KH_2PO_4$ , 12  $NaHCO_3$ , 1.25  $CaCl_2$ , 1.2  $MgCl_2$  (pH 7.2 with  $O_2:CO_2$  95:5, 37 °C). After several minutes incubation the perfusion medium was switched to one containing the same reagents but zero  $Mg^{2+}$ . The flow was about 7 ml  $min^{-1}$  without recirculation (open system). After 10 min the effluent perfusate was collected and  $Mg^{2+}$  was measured by atomic absorbance spectrophotometry (Varian AA-575).  $Mg^{2+}$  (5–20  $\mu M$ ) was present as contaminant in the  $Mg^{2+}$ -free buffer (determined by atomic absorbance spectrophotometry).

times per min did not increase  $Mg^{2+}$  efflux above baseline. Under those conditions the addition of NA induced a  $Mg^{2+}$  efflux similar to that seen in Fig. 1b. Addition of 30 mM KCl or 2 mM caffeine to the perfusing medium also resulted in no detectable  $Mg^{2+}$  efflux above baseline. In separate controls, lactate dehydrogenase and  $K^+$  were measured in the perfusate following the addition of NA (data not shown). The lack of an increase of either in the perfusate indicates that  $Mg^{2+}$  efflux was specific and not due to cell death or a nonspecific increase in cell permeability.

The effect of  $\beta$ -adrenergic stimulation on  $Mg^{2+}$  efflux in perfused hearts is difficult to quantify not only because of cell heterogeneity, but because of changes in heart performance, perfusion rates and other metabolic parameters.

This limitation is overcome by the experiment of Fig. 3 which shows that  $Mg^{2+}$  transport induced by NA is also present in isolated purified rat ventricular myocytes. The addition of NA to a myocyte suspension stimulates  $Mg^{2+}$  efflux in a time- and concentration-dependent manner similar to that observed in perfused hearts. Similarly, the rate of efflux depended on NA concentrations over the range 0.04–10  $\mu M$ . Under these conditions, after 1 min of  $\beta$ -adrenergic stimulation, cells showed a

TABLE 1 Effect of noradrenaline, forskolin, permeant cAMP analogues and carbachol on  $Mg^{2+}$  fluxes in isolated ventricular myocytes

| Agent                                                   | Incubation time (min after addition) |                  |                  |
|---------------------------------------------------------|--------------------------------------|------------------|------------------|
|                                                         | 1                                    | 3                | 5                |
| Extracellular $Mg^{2+}$ content (nmol per $10^6$ cells) |                                      |                  |                  |
| Control                                                 | 40.36 $\pm$ 0.73                     | 40.48 $\pm$ 0.89 | 40.74 $\pm$ 0.85 |
| 10 $\mu M$ NA                                           | 48.52 $\pm$ 1.41                     | 59.48 $\pm$ 1.32 | 65.22 $\pm$ 1.37 |
| 50 $\mu M$ Forskolin                                    | 48.40 $\pm$ 1.87                     | 58.41 $\pm$ 3.42 | 65.50 $\pm$ 1.97 |
| 100 $\mu M$ dBucAMP                                     | 51.24 $\pm$ 1.68                     | 59.56 $\pm$ 0.94 | 63.30 $\pm$ 1.20 |
| 100 $\mu M$ 8-Cl-cAMP                                   | 50.38 $\pm$ 2.12                     | 58.88 $\pm$ 3.40 | 61.40 $\pm$ 1.36 |
| 250 $\mu M$ 8-Br-cAMP                                   | 49.78 $\pm$ 1.56                     | 59.00 $\pm$ 2.32 | 63.52 $\pm$ 3.20 |
| 100 $\mu M$ carbachol                                   | 40.16 $\pm$ 1.29                     | 33.06 $\pm$ 3.64 | 24.84 $\pm$ 0.16 |

Data are means  $\pm$  s.e.m. of 10 experiments for control and NA-stimulated cells and of three experiments for forskolin, cAMP analogues and carbachol-stimulated cells, respectively. Experimental conditions were similar to those described in Fig. 3. dBucAMP,  $N^6,2'$ -O-dibutyladenosine 3':5'-cyclic monophosphate; 8-Cl-cAMP, 8-(4-chlorophenylthio)-adenosine 3':5'-cyclic monophosphate; 8-Br-cAMP, 8-bromoadenosine 3':5'-cyclic monophosphate (all from Sigma).

net release of about 10 nmol  $Mg^{2+}$  per  $10^6$  cells over the normal steady-state level present in the control cells. By prolonging the incubation time, the stimulated cells in the plateau phase released about 25 nmol  $Mg^{2+}$  per  $10^6$  cells before the efflux rate gradually diminished (Fig. 3). Propranolol and sotalol (whose potency is about one-third that of propranolol)<sup>22,23</sup> in doses up to 10 and 30  $\mu M$ , respectively, completely inhibited NA-stimulated  $Mg^{2+}$  extrusion (Fig. 3). Stimulation of  $Mg^{2+}$  efflux was not due to a secondary increase in the fraction of beating cells following NA addition. The addition of 2 mM caffeine (not shown) increased the percentage of cells spontaneously beating to 83% but little or no stimulation of  $Mg^{2+}$  efflux could be observed.

Table 1 conclusively shows that NA stimulates  $Mg^{2+}$  efflux through increasing cardiac cyclic AMP.  $Mg^{2+}$  increases to roughly twice the control value in the extracellular medium 5 min after addition of NA. A similar  $Mg^{2+}$  efflux was observed on incubation of the cells with 50  $\mu M$  forskolin. This concentration of forskolin, by directly activating adenylate cyclase, increases cAMP in myocytes in a quantitatively similar way to the action of NA<sup>24</sup>. Furthermore, identical  $Mg^{2+}$  efflux is observable in the presence of three different permeable cAMP analogues. In contrast, addition of carbachol, which significantly decreases cAMP levels<sup>25</sup>, results in an uptake rather than an efflux of  $Mg^{2+}$  in myocytes.

Taken together, these data indicate a novel hormonally controlled mechanism(s) of  $Mg^{2+}$  transport in cardiac cells. In both perfused rat hearts and isolated myocytes,  $\beta$ -adrenergic stimula-

tion and the resulting rise in cellular cAMP concentrations, activate  $Mg^{2+}$  efflux across the sarcolemma. The rate and amount of efflux are considerable. Measurement of the  $Mg^{2+}$  contents of myocytes by atomic absorbance spectrophotometry in nitric acid extracts of cells sedimented through oil, gave a total  $Mg^{2+}$  content of 110 nmol per  $10^6$  cells. This corresponds to an intracellular concentration of 10.5–11.5 mM, assuming all  $Mg^{2+}$  to be free in solution. Hence, ~20% of the total cell magnesium is mobilized by NA within 5 min. The NA-induced  $Mg^{2+}$  efflux from perfused hearts was between two and three times less than that from isolated myocytes. This difference may be accounted for by perfusion heterogeneity or by a difference in pools of extracellular fluid in cells *in situ* and after isolation. It is not established whether this efflux results from a decrease of cytosolic  $Mg^{2+}$  or from a depletion of  $Mg^{2+}$  stored in intracellular organelles, or both.

All measurements reported here were carried out by atomic absorbance spectrophotometry. This is by far the most sensitive and quantitative method yet available for measuring net magnesium transport. Under our conditions, however, the sensitivity

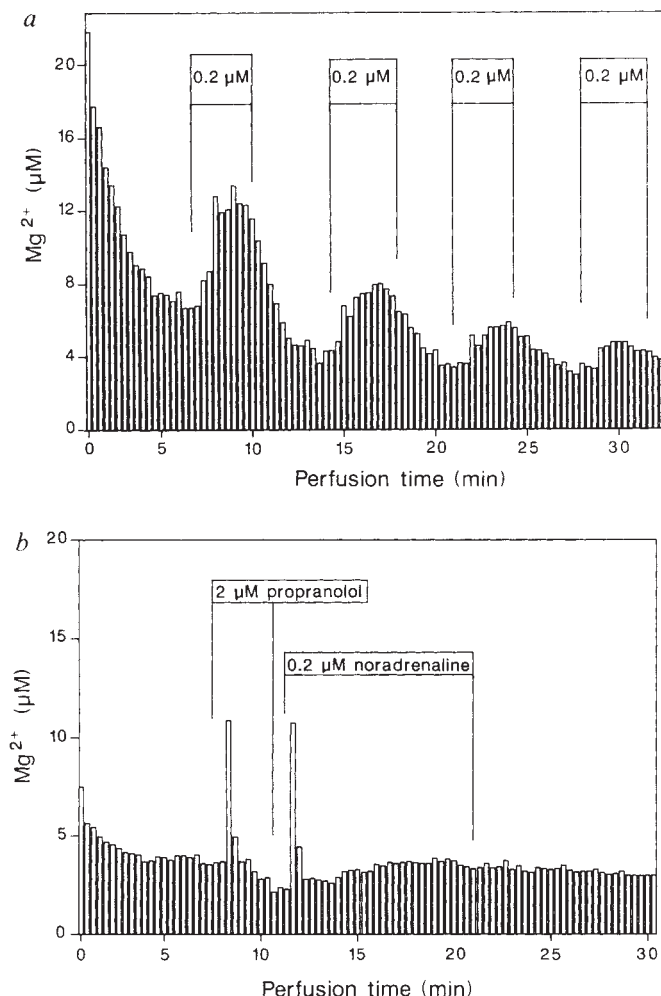


FIG. 2 a,  $Mg^{2+}$  efflux induced by successive NA additions in perfused rat hearts. The buffer was similar to that used in Fig. 1. Where indicated, 0.2 μM NA was directly added in the perfusing medium. Each addition was followed by a resting time in which the heart was washed with the  $Mg^{2+}$ -free buffer previously described in legend to Fig. 1. One experiment typical of four is shown. Experimental conditions were similar to those described in Fig. 1 legend. b, Inhibitory effect of  $\beta$ -blocking agent on the  $Mg^{2+}$  efflux in rat heart. Where indicated, 2 μM propranolol (Sigma) (a  $\beta$ -nonselective antagonist) and 0.2 μM NA were added to the perfusion medium. One experiment typical of three is shown. Experimental conditions were similar to those described in Fig. 1 legend.

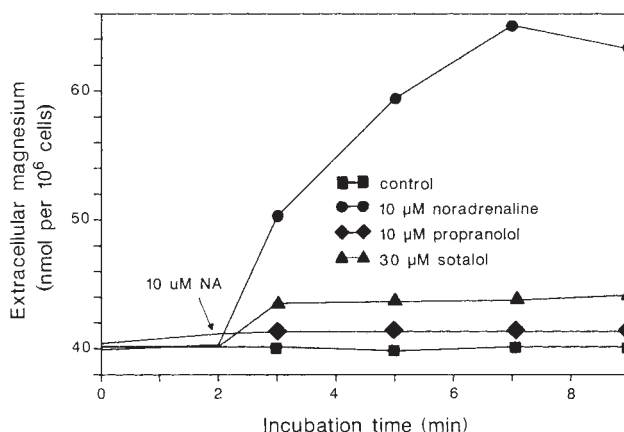


FIG. 3 Stimulatory effect of NA and inhibitory effect of  $\beta$ -adrenergic receptor blocking agents on the  $Mg^{2+}$  efflux in collagenase isolated myocytes. Where indicated, 10 μM NA as agonist and 10 μM propranolol or 30 μM sotalol (Bristol Myers) as  $\beta$ -nonselective antagonists were added. One experiment typical of two is shown.

**METHODS.** Collagenase isolated ventricular myocytes were prepared according to De Young *et al.*<sup>26</sup> and used up to 6 h following isolation. Dispersed myocytes were incubated at 50 μg protein per ml in the  $Mg^{2+}$ -free buffer described in Fig. 1 legend. After stabilization for 5 min, the adrenergic agonist or antagonist was added in the system and at selected times samples were withdrawn and sedimented in microfuge tubes.  $Mg^{2+}$  present in the supernatant was measured by atomic absorbance spectrophotometry. Protein was determined according to Bradford<sup>27</sup>. The percentage of cells spontaneously beating in the control solution was  $20 \pm 3$  ( $n=3$ ) compared with  $78 \pm 4\%$  ( $n=3$ ) in NA.

of the measurement required minimal concentrations of  $Mg^{2+}$  in the perfusate or in the extracellular fluid in which myocytes were suspended. This condition probably results in an amplification of the  $Mg^{2+}$  efflux. Under physiological conditions the concentrations of free  $Mg^{2+}$  inside and outside the cell are similar, whereas our conditions of low extracellular  $Mg^{2+}$  may favour  $Mg^{2+}$  efflux down concentration gradients. It is noteworthy that carbachol, presumably by decreasing cellular cAMP<sup>25</sup>, stimulates  $Mg^{2+}$  uptake across the sarcolemma against the concentration gradient (Table 1).

Our data therefore seem consistent with several mechanisms where high cellular cAMP stimulates  $Mg^{2+}$  efflux and/or low cAMP stimulates  $Mg^{2+}$  uptake. The simplest explanation is that cAMP greatly stimulates magnesium efflux down concentration gradients. More probably, cellular  $Mg^{2+}$  homeostasis could be the result of a 'pump and leak' through one or several transport mechanism(s). In this case cAMP could directly inhibit  $Mg^{2+}$  uptake and/or transiently increase the concentration of cytosolic  $Mg^{2+}$ , and its efflux, through redistribution of magnesium from cellular compartments.

The large  $Mg^{2+}$  fluxes could lead to a significant redistribution of  $Mg^{2+}$  across cells and/or cell compartments which would undoubtedly have a considerable effect on cardiac cell structure and function. For instance, large changes in cytosolic  $Mg^{2+}$  will change the activities of several kinases and ATPases by varying the ratio between free and magnesium-complexed phosphonucleotides, will modulate  $Ca^{2+}$  transport across the sarcolemma, sarcoplasmic reticulum and mitochondria, will change substrate choice and oxidative phosphorylation in mitochondria and will affect key enzyme reactions and metabolic pathways. Finally, an additional role for  $\beta$ -adrenergic response and modulation of cAMP could be visualized in hearts. □

Received 28 February; accepted 8 June 1990.

1. Grubs, R. D. & Maguire, M. E. *Magnesium* **6**, 113–117 (1987).
2. Walker, G. M. *Magnesium* **5**, 9–23 (1986).
3. Flatman, P. W. *J. Memb. Biol.* **80**, 1–14 (1984).
4. Rubin, H. *J. cell. Physiol.* **89**, 613–626 (1977).



5. Ebel, H. & Gunther, T. *J. clin. Chem. clin. Biochem.* **18**, 257-270 (1980).
6. Askari, A., Huang, W. H. & McCormick, P. W. *J. biol. Chem.* **258**, 3453-3460 (1983).
7. Gunther, T. *Magnesium* **5**, 53-59 (1986).
8. Maguire, M. E. *Trends pharmacol. Sci.* **5**, 73-77 (1984).
9. Garlid, K. D. *J. biol. Chem.* **255**, 11273-11279 (1980).
10. Corkey, B. E., Duszynski, J., Rich, T. L., Matschinsky, B. & Williamson, J. R. *J. biol. Chem.* **261**, 2567-2574 (1986).
11. Bond, M., Vadasz, G., Somlyo, A. V. & Somlyo, A. P. *J. biol. Chem.* **262**, 15630-15636 (1987).
12. White, R. E. & Hartzell, H. C. *Science* **239**, 778-780 (1988).
13. Meissner, G. & Henderson, J. S. *J. biol. Chem.* **262**, 3065-3073 (1987).
14. Garfinkel, L. & Garfinkel, D. *Biochemistry* **23**, 3547-3552 (1984).
15. Scarpa, A. & Brinley, F. J. *Fedn Proc.* **40**, 2646-2652 (1981).
16. Lusk, J. E., Williams, J. P. R. & Kennedy, E. P. *J. biol. Chem.* **243**, 2618-2624 (1968).
17. Jasper, P. & Silver, S. *J. Bact.* **133**, 1323-1328 (1978).
18. Osnes, J.-B., Refsum, H., Skomedal, T. & Oye, I. *Acta pharmacol. toxicol.* **42**, 235-247 (1978).
19. Molinoff, P. B. *Drug* **28** (Suppl. 2) 1-15 (1984).
20. Exton, J. H. *Trends pharmacol. Sci.* **3**, 111-115 (1982).
21. Bylund, D. B. *Trends pharmacol. Sci.* **9**, 356-361 (1988).
22. Weiner, N. in *The Pharmacological Basis of Therapeutics* (eds Goodman Gilman, A., Goodman, S. L., Rall, T. W. & Murad, F.) 181-214 (Macmillan, New York, 1985).
23. Waal-Manning, H. *J. Drugs* **12**, 412-441 (1976).
24. Zheng, J.-S., De Young, M. B., Wiener, E., Levy, M. N. & Scarpa, A. *Ann. N.Y. Acad. Sci.* (in the press) (1990).
25. Jakobs, K. H., Aktories, K. & Schultz, G. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **310**, 113-119 (1979).
26. De Young, M. B., Giannattasio, B. & Scarpa, A. *Meth. Enzym.* **173**, 662-676 (1989).
27. Bradford, M. M. *Biochemistry* **72**, 248-254 (1976).

## Loss of pancreatic islet tolerance induced by $\beta$ -cell expression of interferon- $\gamma$

Nora Sarvetnick<sup>\*†</sup>, Judith Shizuru<sup>‡</sup>, Denny Liggitt<sup>\*§</sup>,  
Laura Martin<sup>\*</sup>, Burly McIntyre<sup>‡</sup>, Anita Gregory<sup>‡</sup>,  
Tristram Parslow<sup>||</sup> & Timothy Stewart<sup>\*</sup>

<sup>\*</sup> Department of Developmental Biology, Genentech, Inc.,  
South San Francisco, California 94080, USA

<sup>‡</sup> Department of Immunology, Stanford University, Palo Alto,  
California 94305, USA

<sup>||</sup> Department of Pathology, University of California, San Francisco,  
California 94143, USA

INTERFERON- $\gamma$  (IFN- $\gamma$ ) is produced during the response to infection and participates in immunostimulatory events. We have previously reported the induction of diabetes in transgenic mice (*ins-IFN- $\gamma$* ) in which the expression of the lymphokine IFN- $\gamma$  is directed by the insulin promoter<sup>1</sup>. This diabetes is a result of the progressive destruction of pancreatic islets that occurs with the influx of inflammatory cells. Here we demonstrate that this islet cell loss is mediated by lymphocytes, that engrafted histocompatible islets are destroyed, and that lymphocytes from the transgenic mice are cytotoxic to normal islets *in vitro*. These results indicate that the pancreatic expression of IFN- $\gamma$  can result in a loss of tolerance to normal islets, consistent with its role as an inducer of costimulatory activity<sup>2,3</sup>, which is essential for lymphocyte activation during an immune response.

Sequential histological examination of pancreata from pre-diabetic *ins-IFN- $\gamma$*  transgenic mice reveals that lymphocytes begin to infiltrate at about three weeks old, and is initially characterized by insulitis and peri-insulitis (Fig. 1b). The immune cell infiltrate progresses into widespread inflammation of the pancreas by 10-20 weeks, when disruption of the islet integrity and  $\beta$ -cell loss become apparent and clinical diabetes develops. To determine whether islet destruction in the transgenic mice is caused by infiltrating lymphocytes or through deleterious effects of IFN- $\gamma$ , we backcrossed the *ins-IFN- $\gamma$*  transgene twice onto the severe combined immune deficiency (SCID) strain. SCID mice are homozygous for a recessive allele

(*scid*), that renders them unable to produce mature T and B lymphocytes<sup>4</sup>. In four transgenic, second-backcross progeny lacking mature lymphocytes (that is, SCID) in the spleen, islets were devoid of inflammatory cells and showed no signs of necrosis (Fig. 1c), indicating that infiltrating lymphocytes are primarily responsible for the destruction of the islet cells. There is a mild hyperglycaemia in the *ins-IFN- $\gamma$ /scid* mice, however, with blood glucose levels elevated relative to normal and SCID mice, but not to the high levels found in immunocompetent *ins-IFN- $\gamma$*  transgenic mice, indicating that the transgenic islets are not functioning fully. We conclude that although IFN- $\gamma$  does not itself destroy islet cells, it can damage islet responsiveness to glucose, however the complete destruction of islet cells in the transgenic mice is mediated by lymphocytes, and is not

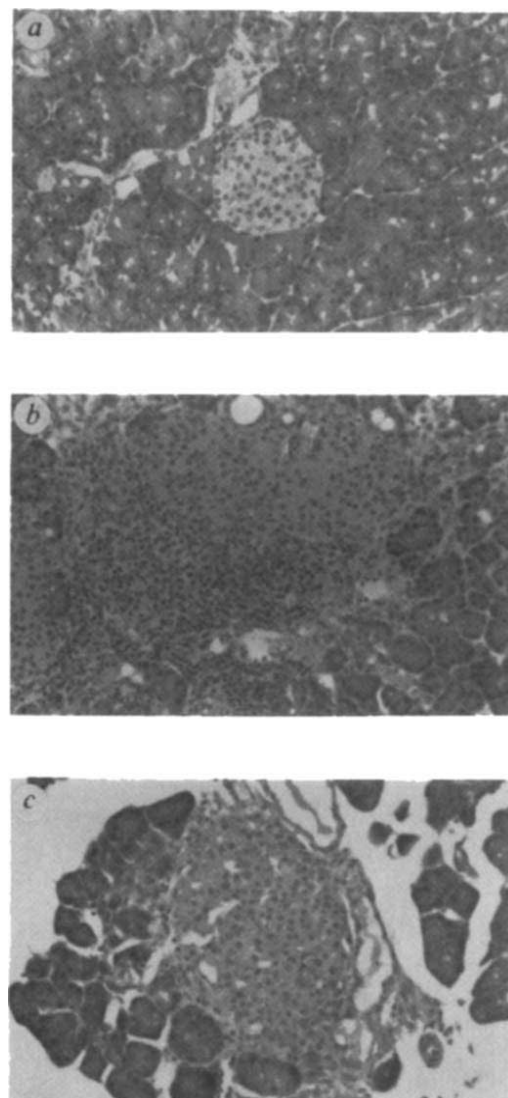


FIG. 1 Histopathology of pancreatic islet in *ins-IFN- $\gamma$*  transgenic mice. *a*, Section of a BALB/c pancreas demonstrating a typical normal islet ( $\times 240$ ). *b*, Section of an *ins-IFN- $\gamma$*  transgenic pancreas from a five-week-old animal. The islet structure is obscured by inflammatory cells ( $\times 240$ ). *c*, Section of an *ins-IFN- $\gamma$ /scid* eight-week transgenic pancreas ( $\times 240$ ). The holey appearance of the *ins-IFN- $\gamma$ /scid* islets is due to the vascular hypertrophy and dilatation.

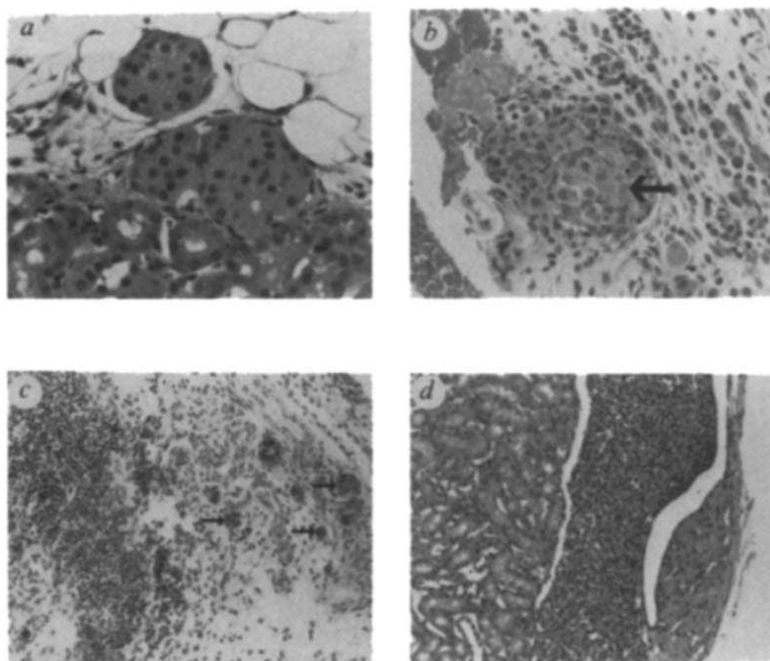
METHODS. Pancreata were fixed in formalin, embedded in paraffin and stained with haematoxylin and eosin for histology. Two independent lines of diabetes-prone *ins-IFN- $\gamma$*  transgenics (461-2 and 462-4) have been analysed<sup>1</sup>; the histopathology of the two lines is virtually identical, but the 461-2 mice are affected earlier than the 462-4 mice and the diabetes is more pronounced in males. Our experiments are on the 461-2 line.

<sup>§</sup> Present address: Department of Comparative Medicine, University of Washington, Seattle, Washington 98195, USA.

<sup>†</sup> To whom correspondence should be addressed at the Department of Neuropharmacology, Scripps Clinic and Research Foundation, 10666 North Torrey Pines Road, La Jolla, California 92037, USA.

FIG. 2 Histopathology of kidneys containing transplanted fetal pancreata. *a*, BALB/c fetal islets grafted into the kidney of BALB/c control recipient, at six weeks after grafting. Subcapsular and perirenal islet cell clusters are free from inflammatory cell involvement ( $\times 240$ ). *b*, BALB/c fetal islet grafted into the kidney of an *ins-IFN- $\gamma$*  recipient, shown two weeks after grafting. The transplanted islet (arrow) is surrounded by loosely arranged granulation tissue containing many mononuclear inflammatory cells. The periphery is infiltrated by inflammatory cells. Individual islet cells are necrotic ( $\times 240$ ). *c*, BALB/c fetal islet grafted into the kidney of an *ins-IFN- $\gamma$*  recipient, six weeks after graft. Marked mononuclear cell accumulation in perirenal tissues. Next to this are several immunoperoxidase-positive islet remnants (arrows) surrounded by connective tissue and inflammatory cells. Note the dark-staining of the cytoplasm of the endocrine cells, which is normally quite pale (see Fig. 1*a*) ( $\times 100$ ). *d*, Adult BALB/c pituitary grafted into the kidney of an *ins-IFN- $\gamma$*  recipient, six weeks after grafting. A portion of a pituitary gland has been transplanted into the subcapsular space. Notice the absence of inflammatory cells ( $\times 100$ ).

**METHODS.** Late-gestational BALB/c females were killed and day-18 embryonic pancreata collected. A single fetal pancreas was implanted into the kidney of each recipient. A month after surgery, the implanted kidney was excised, fixed in formalin and embedded in paraffin. The tissue was then sectioned in 100- $\mu$ m steps to locate the graft. For each graft site, several sections at different levels were analysed by staining with haematoxylin and eosin. To test for insulin expression, adjacent sections were stained with an antibody that reacts with insulin (Dako) and visualized by the immunoperoxidase technique (Vector ABC) using DAB as a chromogen. Eighteen 3–6-month-old BALB/c and 24 *ins-IFN- $\gamma$*  mice were engrafted in the initial experiments. Of the BALB/c recipients, apparently normal adult islets were seen in 14 of the grafts. In the remaining animals no islet tissue was identifiable histologically, indicating that the grafts were either lost from the site or missed in the step-sectioning. Other than scar tissue, no lymphocytic infiltrate was observed in these grafts. In the *ins-IFN- $\gamma$*  transgenics, islet tissue was seen in 16 of the grafts. In 15 cases, this consisted mainly of isolated insulin-immunoreactive cells surrounded by lymphocytes. About 75% of the grafted islet tissue was infiltrated, although occasional islet remnants were unaffected. In one animal there was a scant response, with a few islet remnants being seen fibrous scar tissue that were decorated by few lymphocytes. In the experiments



with collateral adult pituitary/fetal pancreas grafts, we analysed 9 surviving mice of the 12 implanted. Five mice had an identifiable pituitary and pancreas. Each pituitary graft was devoid of lymphocytic infiltration. The founder *ins-IFN- $\gamma$*  transgenic mouse was a CD1  $\times$  BALB/c F1; the line was backcrossed to BALB/c. The transgenic graft recipients were (CD1  $\times$  BALB/c)  $\times$  BALB/c second- and third-backcross animals. These should be histocompatible with the BALB/c grafts. In one experiment, fetal islet tissue was grafted to entire litters of transgenic and non-transgenic *ins-IFN- $\gamma$*  mice. Three litters of 15 animals were engrafted. Nine of these recipients were non-transgenic with six harbouring histologically distinguishable islet grafts. No response was seen in the non-transgenic littermates. This indicates that, although animals are not genetically identical, the CD1 component in these animals does not render the mouse responsive to the graft.

due to the presence of IFN- $\gamma$  *per se*. It is interesting that the islets of the *ins-IFN- $\gamma$ /scid* mice also showed vascular dilatation, which altered their appearance (Fig. 1*c*). Interferon- $\gamma$  is reported to affect endothelial cells *in vitro*<sup>5</sup>, so these changes may be due to direct effects of IFN- $\gamma$  on the vascular endothelium.

To determine whether tolerance to normal non-transgenic islets had been lost in the *ins-IFN- $\gamma$*  mice *in vivo*, we performed histocompatible pancreatic grafts into diabetic and pre-diabetic animals, and evaluated the grafts histologically and/or functionally for signs of rejection. We first surgically implanted histocompatible 18-day fetal pancreata beneath the kidney capsule of the *ins-IFN- $\gamma$*  transgenic mice and their non-transgenic littermates. In non-transgenic recipients, the histological sections revealed apparently normal islets grafted within the periphery of the kidney as expected<sup>6</sup> (Fig. 2). By contrast, in transgenic recipients islet grafts were surrounded by, or infiltrated by lymphocytes, histiocytes, plasma cells and fibroblasts to varying degrees. Subsequent immunochemical staining with an anti-insulin antibody revealed that there were a few scattered insulin-producing cells within most grafts, but the islet structures had become disrupted by inflammatory cells. As a control, we also implanted another histocompatible endocrine tissue, the pituitary, beneath the contralateral kidney capsule. In these grafts, intact pituitary tissue could be identified in both transgenic and non-transgenic animals, with no evidence of rejection. Thus, lymphocytes specific for islet antigens seem to have been activated in these mice.

We also isolated adult histocompatible islets and infused them into the portal vein of the transgenic mice, using *ins-I-A<sup>d</sup>* transgenic diabetic mice<sup>1</sup> and BALB/c mice as controls. Grafted animals were killed and their livers sectioned to visualize the transplanted islets. Histological examination of the grafts in the *ins-IFN- $\gamma$*  transgenic animals revealed a cellular infiltrate that surrounded many of the islets and was first detectable 4 days after transplantation. No islet cell necrosis was seen at this stage but by 2.5 weeks post-transplantation islet cell loss was evident, and the scant number of islets remaining had haloes of varying densities made up of inflammatory cells (Fig. 3). No lymphocytic infiltrate was observed in the control *ins-I-A<sup>d</sup>* transgenic recipients or BALB/c non-transgenic mice, confirming that islet-responsive lymphocytes are present in the *ins-IFN- $\gamma$*  transgenic mice.

In chemically induced diabetes, the portal vein infusion of islets brings about implantation of enough islets in the liver to reverse hyperglycaemia in the diabetic recipients<sup>7</sup>. Within 24 hours of surgery the blood glucose levels of our control *ins-I-A<sup>d</sup>* animals had returned to normal, but did not in the case of the *ins-IFN- $\gamma$*  transgenic mice, all of which had blood glucose concentrations of  $>400$  mg dl<sup>-1</sup> before and at 24 hours after surgery. Histochemistry revealed a marked paucity of immunoreactive-insulin in the transplanted islets in the *ins-IFN- $\gamma$*  transgenic mouse livers at two and twenty days after grafting, compared with *ins-I-A<sup>d</sup>* transgenic mice. The expression of other islet hormones, glucagon and somatostatin, was normal as determined by immunochemistry.



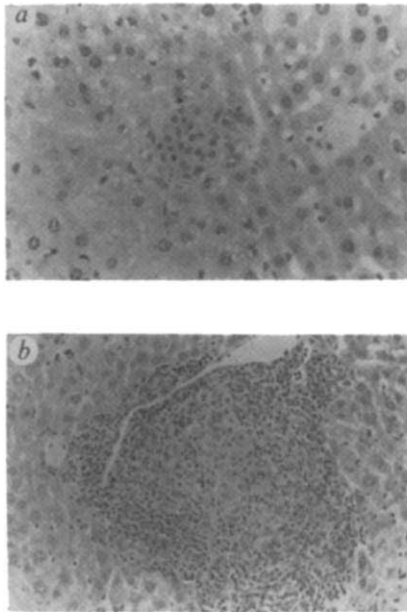


FIG. 3 Histopathology of livers containing transplanted adult islets. *a*, BALB/c islet within the liver of *ins-I-A<sup>d</sup>* transgenic. Islet (arrow) transplanted into liver lacks and inflammatory cell component ( $\times 240$ ). *b*, BALB/c islet in the liver of an *ins-IFN- $\gamma$*  transgenic mouse. There is a marked inflammatory cell response seen surrounding fragmented transplanted islet cells ( $\times 240$ ).

**METHODS.** Adult BALB/c islets were isolated by the collagenase perfusion method. The common bile duct was cannulated and the pancreas infiltrated with collagenase ( $0.8 \text{ mg ml}^{-1}$ ) in HBSS (Hanks buffered saline solution). The pancreas was excised and incubated at  $37^\circ \text{C}$  for 25 min. After disruption by pipetting the tissue was washed by centrifuging twice and then passed through a strainer. Islets were separated by centrifugation through a Ficoll zone gradient. Banded islets were washed and incubated in DME containing 10% fetal calf serum. Eight hundred islets were infused into the portal vein of each recipient. Six diabetic *ins-IFN- $\gamma$*  recipients and two diabetic *ins-I-A<sup>d</sup>* recipients were infused with the isolated islets. The *ins-I-A<sup>d</sup>* recipients were killed one month after surgery. The *ins-IFN- $\gamma$*  recipients were analysed at one day (2 mice), four days (2 mice), and two-and-a-half weeks (2 mice) after surgery. Livers were fixed in formalin and embedded in paraffin. Sections were cut and stained with haematoxylin and eosin. Adjacent sections were also analysed for insulin, glucagon and somatostatin expression by immunohistochemistry (not shown).

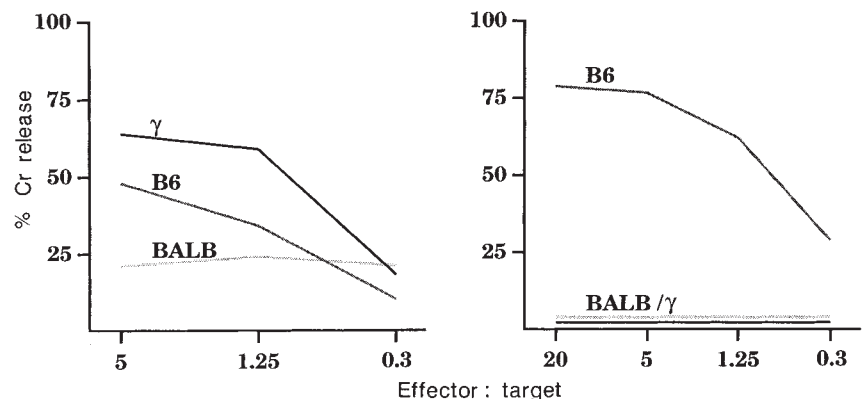
Our acute failure to reverse diabetes in *ins-IFN- $\gamma$*  mice probably reflects an inability of the hepatic islet grafts to produce sufficient insulin. One interesting possibility is that IFN- $\gamma$  produced in the transgenic mice is inhibiting the function of the transplanted islets. Mildly deleterious effects of high levels of IFN- $\gamma$  ( $2,000 \text{ units ml}^{-1}$ ) on pancreatic islets *in vitro* have been observed<sup>8</sup>. Although we have not been able to detect any IFN- $\gamma$ -like antiviral activity in serum derived from the tail vein of our transgenic mice, both the impaired islet function in the *ins-IFN- $\gamma$ /scid* mice and the proximity of the hepatic circulation to the pancreas suggest that local circulating levels of this lymphokine might be responsible for the failure of the grafts to reverse diabetes in the transgenic mice.

The pancreatic and islet graft experiments demonstrate that lymphoid cells in the transgenic mice can respond to histocompatible, non-transgenic islet grafts. To ascertain whether this reflects immunological sensitization to normal islets in the trans-

genic mice rather than IFN- $\gamma$ -mediated damage, we investigated whether islet-specific cytotoxic lymphocytes could be found in the *ins-IFN- $\gamma$*  transgenic mice. In cytotoxicity experiments we used chromium-loaded islets as targets and lymph node cells derived from the transgenic mice as effectors. Lymph node cells from transgenic and non-transgenic mice were primed with either irradiated islets or spleen cells and then incubated with either chromium-loaded islet or spleen cells as targets. This *in vitro* experiment is therefore independent of the potential effects of circulating lymphokines. Our results are presented in Fig. 4. The primed lymph node cells from the transgenic mice induced chromium release when histocompatible islets but not histocompatible spleen cells were used as targets. Therefore lymphocytes isolated from the *ins-IFN- $\gamma$*  transgenic mice can be cytotoxic to normal islets. When protease-treated dispersed islet cells were loaded with chromium and used as targets, we found no cytotoxicity with the *ins-IFN- $\gamma$*  (or the allogeneic C57BL/6)

FIG. 4 Cytotoxicity of lymph node cells from *ins-IFN- $\gamma$*  transgenic, BALB/c, and C57BL/6 (B6) mice to islets and spleen cells *in vitro*. Left, Per cent chromium released when islets are used as targets. The genotype of the effector cells is designated on the graph. Right, Per cent chromium released when concanavalin A-stimulated spleen cells are used as targets.

**METHODS.** The mesenteric lymph nodes adjacent to the pancreas were excised from *ins-IFN- $\gamma$* , BALB/c and C57BL/6 mice. Islet cytotoxicity: To measure cytotoxic responses to BALB/c islets,  $2 \times 10^7$  lymph node cells from *ins-IFN- $\gamma$* , BALB/c, and C57BL/6 mice were each primed with 800 irradiated BALB/c islets and  $2 \times 10^6$  irradiated spleen cells in 10 ml RPMI for 5 days. BALB/c irradiated islets and BALB/c irradiated spleen cells were used to prime the *ins-IFN- $\gamma$*  and BALB/c lymph node cells. BALB/c irradiated islets and C57BL/6 irradiated spleen cells were used to prime the C57BL/6 lymph node cells. After the 5-day incubation, primed lymph node cells were collected and then incubated with chromium-loaded BALB/c whole islets (50) in the ratios shown for 5 hours. For deriving effector:target ratios with whole islets, each islet was assumed to contain an average of 1,000 cells, as determined from trypsin dispersal studies. Each point was performed in triplicate and averaged to produce the graph. The entire experiment was repeated three times. Following incubation, supernatants were counted



in a gamma counter. Spleen cytotoxicity: To measure cytotoxic responses to BALB/c spleen cells,  $2 \times 10^7$  lymph node cells from each group (*ins-IFN- $\gamma$* , BALB/c, C57BL/6) were primed with  $2 \times 10^6$  irradiated BALB/c spleen cells. After the 5-day incubation, the lymph node cells from each group were collected and then incubated with chromium-loaded, BALB/c derived, concanavalin A-treated spleen cells for 5 hours at the ratios given. Each point was performed in triplicate and averaged. Following the incubation, supernatants were counted in a gamma counter.

lymphocytes. This indicates that the observed cytotoxicity depends on the integrity of the islet cell surface constituents.

Taken together, our results show that pancreatic expression of IFN- $\gamma$  can result in immune sensitization of the transgenic mice to islets. In other experiments involving ectopic expression of antigens in islets, no islet immune-related destruction results, irrespective of whether tolerance is induced<sup>9</sup>. Indeed, there are other examples of T cells reactive to components in the periphery

existing in a quiescent state<sup>10-19</sup>. *In vitro* experiments have demonstrated that IFN- $\gamma$  can induce co-stimulatory activity leading to T cell activation<sup>2,3</sup>, so when produced *in vivo*, it might activate quiescent autoreactive cells, stimulating lymphocytes to destroy both transgenic and normal islets. Our work implies that *in vivo* the lymphokine IFN- $\gamma$  activates T cells so that they do not remain anergic to antigens encountered in the periphery. □

Received 9 April; accepted 14 June 1990.

1. Sarvetnick, N., Liggitt, D., Pitts, S. L., Hansen, S. E. & Stewart, T. A. *Cell* **52**, 773-782 (1988).
2. Hawrylowicz, C. M. & Unanue, E. R. *J. Immunol.* **141**, 4083-4088 (1988).
3. Weaver, C. T., Hawrylowicz, C. M. & Unanue, E. R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8181-8185 (1988).
4. Bosma, C. G., Custer, R. P. & Bosma, M. J. *Nature* **301**, 527 (1983).
5. Duijvestijn, A. M., Schreiber, A. B. & Butcher, E. C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9114-9118 (1986).
6. Brown, J., Molnar, I. G., Clark, W. & Mullen, Y. *Science* **184**, 1377-1379 (1974).
7. Kemp, C. V., Knight, M. G., Scharp, D. W., Lacy, P. E. & Ballinger, W. F. *Nature* **244**, 447 (1973).
8. Campbell, I. L., Iscrao, A. & Harrison, L. C. *J. Immunol.* **141**, 2325 (1988).
9. Harrison, L. C., Campbell, I. C., Allison, J. & Miller, J. F. A. P. *Diabetes* **38**, 815-818 (1989).
10. Mitchison, N. A. *Immunology* **15**, 509-530; 531-547 (1968).
11. Hooper, D. C., Young, J. L., Elson, C. J. & Taylor, R. B. *Cell. Immunol.* **106**, 53-61 (1987).

12. Hooper, D. C. & Taylor, R. B. *Eur. J. Immunol.* **17**, 797 (1987).
13. Rammensee, H. G., Kroschewski, R. & Frangoulis, B. *Nature* **339**, 541-544 (1989).
14. Qin, S., Cubbold, S., Benjamin, R. & Waldmann, H. *J. exp. Med.* **169**, 779-794 (1989).
15. Markmann, J. et al. *Nature* **336**, 476-479 (1988).
16. Burki, L. C., Lo, D., Kanagawa, O., Brinster, R. L. & Flavell, R. A. *Nature* **342**, 564-566 (1989).
17. Lo, D., Burki, L. C., Flavell, R. A., Palmer, R. D. & Brinster, R. L. *J. exp. Med.* **170**, 87-104 (1989).
18. Bohme, J. et al. *Science* **211**, 1179-1183 (1989).
19. Murphy, K. M., Weaver, C. T., Elish, M., Allen, P. M. & Loh, D. Y. *Proc. natn. Acad. Sci. U.S.A.* **86**, 10034-10038 (1989).

ACKNOWLEDGEMENTS. This work was supported by an MDA postdoctoral fellowship (N.S.), by a JDF postdoctoral fellowship (J.S.) and an NIH grant (J.S. and B.M.). T.G.P. is a scholar of the Leukemia Society of America. N.S. and J.S. thank Dr Garry Fathman for support.

## Targeted disruption of the murine *int-1* proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development

Kirk R. Thomas & Mario R. Capecchi\*

Howard Hughes Medical Institute, Department of Biology and Human Genetics, Salt Lake City, Utah 84112, USA

THE *int-1* proto-oncogene was first identified as a gene activated in virally induced mouse mammary tumours<sup>1,2</sup>. Expression studies, however, suggest that the normal function of this gene may be in spermatogenesis and in the development of the central nervous system<sup>3-5</sup>. Genes sharing sequence similarity with *int-1* have been found throughout the animal kingdom. For example, *int-1* has 54% amino-acid identity to the *Drosophila* segment polarity gene *wingless* (*wg*)<sup>6</sup>. Both the *int-1* and *wg* gene products seem to be secreted proteins, presumably involved in cell-cell signalling<sup>7-11</sup>. We have now explored the function of *int-1* in the mouse by disrupting one of the two *int-1* alleles in mouse embryo-derived stem cells using positive-negative selection<sup>12</sup>. This cell line was used to generate a chimaeric mouse that transmitted the mutant allele to its progeny<sup>13-16</sup>. Mice heterozygous for the *int-1* null mutation are normal and fertile, whereas mice homozygous for the mutation may exhibit a range of phenotypes from death before birth to survival with severe ataxia. The latter pathology in mice and humans is often associated with defects in the cerebellum. Examination of *int-1*<sup>-</sup>/*int-1*<sup>-</sup> mice at several stages of embryogenesis revealed severe abnormalities in the development of the mesencephalon and metencephalon indicating a prominent role for the *int-1* protein in the induction of the mesencephalon and cerebellum.

Gene targeting has been used to disrupt numerous genes in mouse embryo-derived stem (ES) cells<sup>12,17-28</sup> and in several cases, germ-line transmission of the mutated gene has been reported<sup>25-28</sup>. The targeting vector we used to mutate *int-1*, pINT-1-N/TK, is shown in Fig. 1a. It contains 13 kilobases (kb) of mouse genomic DNA surrounding the *int-1* structural gene, which is interrupted in exon 2 by a 1-kb fragment containing the neomycin resistance (*neo*<sup>r</sup>) gene, and flanked by copies of the thymidine kinase (*TK*) gene from herpes simplex virus.

TABLE 1 Phenotype of *int-1*<sup>-</sup>/*int-1*<sup>-</sup> mice

### (a) Mortality of *int-1*<sup>-</sup>/*int-1*<sup>-</sup> mice

| No. litters | No. pups | No. <i>int-1</i> <sup>-</sup> / <i>int-1</i> <sup>-</sup> | Day of death of <i>int-1</i> <sup>-</sup> / <i>int-1</i> <sup>-</sup> pups |         |     |
|-------------|----------|-----------------------------------------------------------|----------------------------------------------------------------------------|---------|-----|
|             |          |                                                           | 0-0.5                                                                      | 0.5-1.5 | >30 |
| 5           | 42       | 10                                                        | 6                                                                          | 3       | 1   |

### (b) Genotype and phenotype of embryos

| Days p.c. | Genotype | No. individuals | Brain morphology |
|-----------|----------|-----------------|------------------|
| 14.5      | +/+      | 3               | Normal           |
|           | +/-      | 3               | Normal           |
|           | -/-      | 2               | Underdeveloped   |
| 17.5      | +/+      | 7               | Normal           |
|           | +/-      | 7               | Normal           |
|           | -/-      | 4               | Underdeveloped   |

Mice heterozygous at *int-1* were mated. (a) Mortality of live births: Five pregnancies proceeded to term and the nests examined at daily intervals. Dead offspring were removed and their DNA examined as described in the legend to Fig. 2. At day 30, DNA was extracted from tails of all living offspring and analysed by the same method. (b) Genotype and phenotype of embryos. Embryos were surgically removed from pregnant mothers at the times indicated and immersed in Bouin reagent. Yolk sacs were genotyped as described in Fig. 2. After 2 days in fixative, the embryonic brains were surgically removed, examined under a dissecting microscope and qualitatively compared. The 14.5-day embryos were from a single mother, the 17.5 embryos represent the sum of two pregnancies.

The *neo*<sup>r</sup> gene serves both to disrupt the *int-1* protein reading frame and to provide a positive selectable marker, resistance to growth in G418, for the presence of the *int-1* mutation. The *TK* genes, lying outside *int-1* homology, function in the presence of the antiviral agent gancyclovir (GANC), as negative selectable markers that allow enrichment for cells in which homologous recombination has occurred<sup>12</sup>.

ES cells were transformed by electroporation with the vector, pINT-1-N/TK and then grown in the presence of G418 and GANC. DNA was extracted from those colonies resistant to both drugs and analysed by Southern transfer for an insertion mutation at *int-1*. Of  $5 \times 10^8$  cells surviving electroporation,  $5 \times 10^6$  gave rise to G418<sup>r</sup> colonies and 400 were resistant to both G418 and GANC. One of the G418<sup>r</sup>-GANC<sup>r</sup> colonies, cell-line 54, proved to contain a targeted disruption in one of the two *int-1* genes. The targeted disruption frequency for *int-1* was considerably lower than previously reported for *hprt* (ref. 17), *int-2* (ref. 12), *en-2* (ref. 20), *c-abl* (ref. 25), *IGF-II* (ref. 27) and  $\beta_2$ -microglobulin<sup>22,24</sup>, which may reflect inherent differences of the loci.

\* To whom correspondence should be addressed.



The demonstration of gene targeting in cell-line 54 is illustrated in Fig. 1b. DNAs from the parental ES cell line (CC1.2)<sup>13</sup> and cell-line 54 were digested with *Sac*I and analysed by Southern blotting using a hybridization probe (probe A) that contained sequences outside those found on the targeting vector. In the parental line, this probe detected a 4-kb fragment of DNA. In the mutant cell line, an additional *int-1* band of 5 kb, was evident, consistent with the insertion of the 1-kb *neo*<sup>r</sup> gene into *int-1*. This interpretation was verified by digestion of this DNA with two additional restriction enzymes and probing with *neo* sequences (data not shown). The same DNAs were digested with *Bgl*II and *Xho*I and analysed with a probe (probe B) contained within the vector sequences. In addition to the wild-type allele, DNA from cell-line 54 also revealed two fragments derived from the mutant, *int-1*<sup>-</sup> allele. One of these, of 4 kb, is that predicted for the *neo*<sup>r</sup> insertion; the other, of 4.3 kb, is from a tandem duplication (described in the legend to Fig. 1) and subsequent insertion of vector sequences at the target site. This duplication-insertion configuration was verified by six independent enzyme digests (data not shown) and is similar to targeting events described by others<sup>29,30</sup>.

Chimaeric mice were generated by microinjecting the male ES cell-line 54—originally derived from a 129 Sv agouti mouse—into blastocysts from C57Bl/6 nonagouti mice. One male chimaera, when mated to a C57Bl/6 female sired pure agouti offspring indicating that his germ line contained sperm derived from ES cell-line 54. DNA prepared from tail biopsies taken from the chimaera and two of his agouti offspring was analysed by Southern transfer as shown in Fig. 2a. DNA from the parental ES cell line and cell-line 54 serve as controls for the *int-1*<sup>+</sup>/*int-1*<sup>+</sup> and *int-1*<sup>-</sup>/*int-1*<sup>+</sup> genotypes, respectively. As expected, the chimaeric father contained *int-1*<sup>-</sup> sequences, but not in stoichiometric amounts, a reflection of the degree of chimaerism

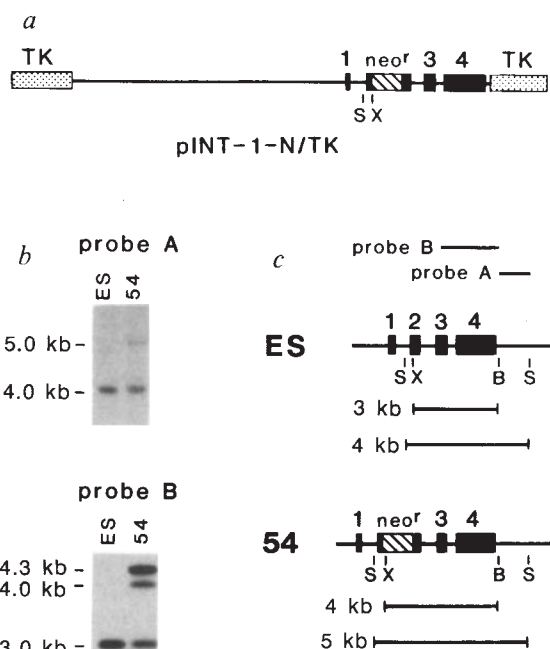
in the tail. Both of his offspring, however, show the same *int-1*<sup>-</sup>/*int-1*<sup>+</sup> genotype stoichiometry seen in the original mutant cell line.

The heterozygous offspring were bred with C57Bl/6 mice to establish a colony of mice heterozygous for the *int-1*<sup>-</sup> mutation. Sibling heterozygotes were mated and the pregnancies allowed to proceed to term. All offspring were genotyped at the *int-1* locus by analysing tail DNA by Southern blotting. Table 1 summarizes the fate of ten *int-1*<sup>-</sup>/*int-1*<sup>-</sup> offspring (from five litters): six were dead within 0.5 days of birth; three survived 0.5 days but died before day 1.5 and one survived to adulthood. Variation in survival may reflect variation in the penetrance of the *int-1* mutation or variability in the nurturing by the mother; female mice will often separate malfunctioning offspring from the rest of the litter. But in our experiments, one out of six homozygous *int-1*<sup>-</sup> mice (two litters) examined before birth at 17.5 days p.c. died *in utero*.

The *int-1*<sup>-</sup>/*int-1*<sup>-</sup> mouse that survived to adulthood was most informative regarding *int-1* function. She exhibited severe ataxia. When moving, she pivoted both clockwise and counter-clockwise around her rear legs. This abnormal pattern of movement was not due to muscle wastage or defects in local motor neuron circuitry as she could jump more than 30-cm vertically. Relative to heterozygous siblings, she responded normally to stimuli of light, sound, smell and touch.

The presence of this behavioural phenotype in the *int-1*<sup>-</sup> homozygote (that is, the loss of balance and coordinated movement) suggested a defect in the cerebellum and prompted us to examine the central nervous system (CNS) in embryos produced from heterozygous matings. Embryos were removed from pregnant females at 11.5, 14.5 and 17.5 days p.c. The DNA was extracted from the yolk sacs and analysed by Southern transfer to determine the *int-1* genotype; the embryos were then fixed

FIG. 1 a, The targeting vector, pINT-1-N/TK, consists of 13.5 kb of mouse genomic sequences, containing the *int-1* gene interrupted at the *Xho*I site in exon 2 (ref. 2) with the *neo*<sup>r</sup> gene from pMC1-Neo (ref. 17). The *int-1*-*neo*<sup>r</sup> sequences are flanked by copies of the *HSV-TK* gene from pMC1-TK (ref. 12). Narrow lines, introns; black boxes, exons; hatched boxes, *neo*<sup>r</sup>; stippled boxes, *HSV-TK*. The vector was linearized by digestion with *Sac*I, and introduced into ES cell line CC1.2 by electroporation. Aliquots of cells were subjected to three growth conditions: non-selective medium to assess viability (50% of the starting cells survived electroporation); G418 medium to assay the fraction of cells stably transformed by the vector; and G418 plus GANC medium to enrich for transformants containing a *neo*<sup>r</sup> insertion at one of the *int-1* loci. DNA from cells surviving G418-GANC selection were subjected to Southern transfer analysis as shown in b. Note: The presence of flanking *TK* sequences was included to add an additional enrichment to the positive-negative selection protocol. Previous data suggested that a large fraction of non-targeted, yet GANC<sup>r</sup>, cells transformed with vector containing a single *TK* survived because of transfection-induced alterations of the *TK* gene<sup>12</sup>. The presence of the second copy of the *TK* sequences does seem to enhance the enrichment factor from the 10<sup>3</sup> seen previously, to 10<sup>4</sup> (that is, 5×10<sup>6</sup>/4×10<sup>2</sup>), seen in Table 1. We have shown that blocking none, one, or both ends of a targeting vector has no detectable effect on the frequency of homologous recombination at the *hprt* locus (K.R.T. and M.R.C., unpublished observations). b, Identification of an *int-1*<sup>+</sup>/*int-1*<sup>-</sup> cell line. ES, DNA from cell line CC1.2; 54, DNA from the G418<sup>r</sup>, GANC<sup>r</sup> cell line transfected with pINT-1-N/TK. Sizes of the bands are indicated (kb). DNAs hybridized to probe A were digested with *Sac*I; those to probe B with *Bgl*II and *Xho*I. The *int-1*<sup>+</sup> allele gives a stronger signal than the *int-1*<sup>-</sup> allele because of wild-type sequences in the feeder cells used to propagate the ES cells. c, A schematic representation of the Southern transfer data. ES is a map of the wild-type *int-1*<sup>+</sup> locus; 54 is a map of the *int-1*<sup>-</sup> allele containing a *neo*<sup>r</sup> insertion in exon 2. Beneath each gene are shown the positions of diagnostic restriction endonuclease cleavage sites, and the length of the digestion products. Positions of the probes are indicated above the maps. Probe A (DNA sequence not present in the targeting vector) is a 1-kb *Bgl*II-*Hind*III fragment 3' of the *int-1* gene (gift of R. Nusse); Probe B is a 2-kb *Clal*-*Bam*HI fragment within the *int-1* gene. B, *Bgl*II; S, *Sac*I; X, *Xho*I. When DNA from cell line 54 is hybridized with probe B, which consists of sequences from within the vector pINT-1-N/TK, an unexpected fragment 4.3 kb in length is seen. This fragment represents additional copies of the



vector sequences that have integrated in a tandem array 5' to those sequences depicted. Such a configuration is most likely the result of the head-to-tail concatamerization of the vector before its recombination with the target locus. The 4.3-kb fragment extends from the *Xho*I site in exon 2 to a *Bgl*II site in the targeting vector and is predicted from the map of the vector. Digestions of DNA from cell-line 54 with six additional restriction endonucleases and hybridization with *neo*<sup>r</sup> probes verify this configuration. The above analysis also showed that DNA sequences both 5' and 3' to the *int-1* locus were not altered, demonstrating that the targeted mutation was limited to the *int-1* gene.

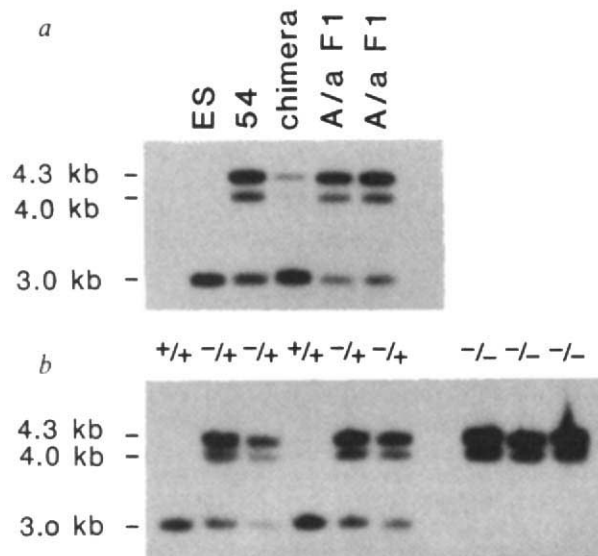


FIG. 2 Transmission of the *int-1*<sup>-</sup> genotype. *a*, Cell lines and founding animals. ES cell line 54 was microinjected into 4.5 day C57Bl/6 preimplantation embryos, surgically transferred into the uterus of pseudopregnant mice and allowed to come to term. Two resulting male chimaeras were test bred with C57Bl/6 females for transmission of agouti offspring. One of these males proved to be a germ-line chimaera. DNA was extracted from cell lines (ES and 54) or from tail biopsies, digested with *Xho*I and *Bgl*II, and analysed with probe B as described in the legend to Fig. 1. ES, the parental cell line CC1.2; 54, the *int-1*<sup>+</sup>/*int-1*<sup>-</sup> cell line carrying the targeted mutation in *int-1*. Chimaera designates the founding chimaeric male; A/a and A/a are his agouti offspring produced following mating with a C57Bl/6 female. *b*, Transmission to embryos. Heterozygous offspring from A/a and A/a shown in *a* were mated. Following 17.5 days gestation, the embryos were removed from the mother. DNA was extracted from the yolk sac surrounding each individual embryo and analysed by Southern transfer as described in Fig. 1. The genotype at *int-1* is indicated above each lane: +, wild type; -, *neo*<sup>r</sup> insertion in exon 2.

and dissected. DNA analyses of nine embryos from one pregnancy is illustrated in Fig. 2*b* and all three expected genotypes were clearly present. The results of visual analyses of the fetal brains are summarized in Table 1(*b*). After 11.5 days of gestation, morphological differences between heterozygous and homozygous *int-1*<sup>-</sup> mice were not readily apparent. But by 14.5 days, malformations in the mesencephalon and metencephalon were apparent. At 17.5 days, this malformation was translated into mild midbrain hydrocephaly. This malformation was 100% concordant with the homozygous *int-1*<sup>-</sup> genotype: 6/6 embryos of *int-1*<sup>-</sup>/*int-1*<sup>-</sup> genotype exhibited this phenotype, whereas 20/20 *int-1*<sup>+</sup>/*int-1*<sup>+</sup> and *int-1*<sup>+</sup>/*int-1*<sup>-</sup> embryos showed normal brain development. Figure 3*a* and *b* compares dorsal views of brains from *int-1*<sup>-</sup>/*int-1*<sup>+</sup> and *int-1*<sup>-</sup>/*int-1*<sup>-</sup> embryos at 17.5 days of gestation. It is apparent from macroscopic examination alone, that the mesencephalon and metencephalon are highly defective. This is corroborated by examination of sagittal and parasagittal sections of the same brains (Fig. 3*c*, *d*). In the *int-1*<sup>-</sup> homozygote, the cerebellum and a large portion of the mesencephalon, as indicated by the arrows, are not visible. As a reference point, note that the posterior choroid plexus (pcp) is present in both brains. The embryos used for the above analysis, from the same mating, showed normal embryonic leg and head movement and touch response. The telencephalon as judged from examination of both parasagittal and coronal sections, appeared to be normal. However, in the *int-1*<sup>-</sup> homozygous mouse, the frontal lobes of the brain do seem to extend more caudally relative to the heterozygous mate, perhaps a result of expansion into regions normally occupied by the missing parts of the midbrain.

The *int-1*<sup>-</sup>/*int-1*<sup>-</sup> mouse that survived until adulthood exhibited hydrocephaly in the caudal region of the cerebral hemispheres as well as in the midbrain. Examination of sagittal and parasagittal sections of this *int-1*<sup>-</sup>/*int-1*<sup>-</sup> brain corroborated the presence of the hydrocephaly but also unexpectedly showed that the posterior portion of the cerebellum was present (Fig. 3*e*, *f*). In this context, it is interesting that Wilkinson *et al.*<sup>5</sup> did not observe *int-1* messenger RNA in the 10.5-day rostral hindbrain. Furthermore, LeDouarin and co-workers have recently provided evidence that in the chick, the anterior portion of the cerebellum may be derived from the mesencephalon, whereas the posterior portion comes from the metencephalon<sup>31</sup>. If this is also true for the mouse, then the phenotypic influence of the *int-1* mutation in this mouse may have respected the mesencephalon-metencephalon boundary.

In summary, heterozygous *int-1*<sup>-</sup> mice seem normal. A prominent effect of disrupting both *int-1* alleles is a major defect in the development of the midbrain and cerebellum. In fact, it seems that *int-1* protein is required in the induction of a large portion of the mesencephalon and of the cerebellum as in its absence these structures are not formed normally. Further, the *int-1* homozygote that survived to adulthood displayed the classical, although a more extreme, form of symptomatology attributed to human and mouse cerebellar defects (that is, severe ataxia).

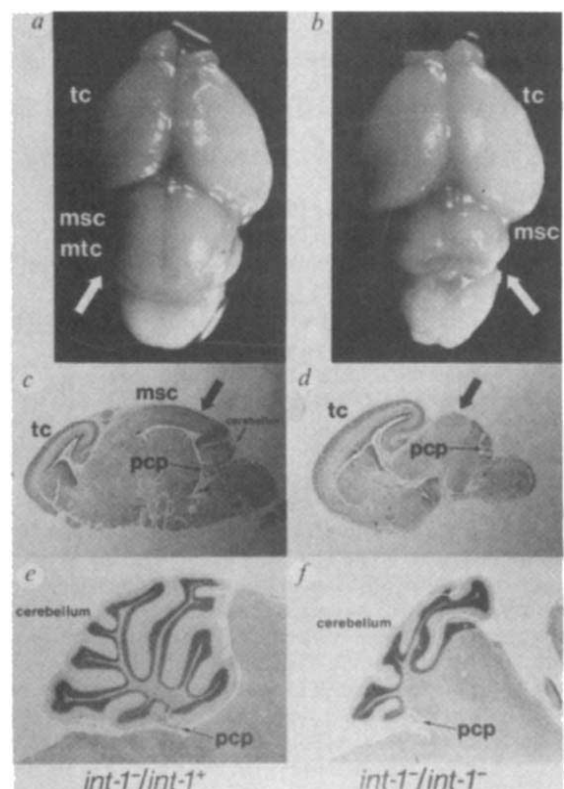


FIG. 3 Comparison of brains from heterozygous and homozygous, *int-1*<sup>-</sup> mice. *a* and *b*, 17.5 day embryos were fixed in Bouin's reagent (Sigma). The yolk sacs were removed for DNA analysis as described in Table 1. Following two days in fixative, the brains were dissected, rinsed in PBS and photographed at  $\times 6$  magnification. The field of view is  $5 \times 10$  mm. Arrows indicate the cerebellar region absent in the homozygote. *c* and *d*, The brains shown in *a* and *b* were embedded in paraffin, sectioned ( $10 \mu\text{m}$ ), and stained by haematoxylin and eosin (H and E) regressive staining. The field of view is  $6 \times 4$  mm. Arrows indicate mesencephalic tissue absent in the homozygote. *e* and *f*, Brains were dissected from adult (5 week) mice, fixed in PLP, embedded in paraffin, sectioned ( $8 \mu\text{m}$ ) stained with H and E. The field of view is  $6 \times 4$  mm. tc, telencephalon; msc, mesencephalon; mtc, metencephalon; pcp, posterior choroid plexus.



In the 17.5-day embryo, cellular structures corresponding to the cerebellum are not observed. But in the mouse that survived to adulthood, the posterior portion of the cerebellum was present. An intriguing hypothesis to explain this difference is that the dependency on the inductive influence of *int-1* exhibits a rostral-caudal gradient beginning in the mesencephalon and spreading to the metencephalon. This could produce differences in the penetrance of the *int-1*<sup>-</sup> mutation such that some mice lack only the anterior portion of the cerebellum, whereas others are missing the entire cerebellum.

The *int-1* gene is expressed during the ontogeny of the CNS along the dorsal midline of the developing neural tube from the midbrain to the tail<sup>5</sup>. From analysis of sagittal and transverse sections, in either embryos or in the *int-1*<sup>-</sup> survivor, we have not observed morphological defects in the spinal cord. This may indicate that either *int-1* expression in this region of the developing CNS is spurious or that another gene, perhaps another member of the *int-1* family, can substitute for *int-1* function in the development of the spinal cord. □

Received 20 July; accepted 6 August 1990.

1. Nusse, R. & Varmus, H. E. *Cell* **31**, 99–109 (1982).
2. van Ooyen, A. & Nusse, R. *Cell* **39**, 233–240 (1984).
3. Jakobovits, A., Shackleford, G. M., Varmus, H. E. & Martin, G. R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7806–7810 (1986).
4. Shackleford, G. M. & Varmus, H. E. *Cell* **50**, 89–95 (1987).
5. Wilkinson, D. G., Bailes, J. A. & McMahon, A. P. *Cell* **50**, 79–88 (1987).
6. Rijsewijk, F. et al. *Cell* **50**, 649–657 (1987).
7. Morata, G. & Lawrence, P. A. *Dev. Biol.* **56**, 227–240 (1977).
8. Baker, N. E. *EMBO J.* **6**, 1765–1770 (1987).
9. Papkoff, J., Brown, A. M. C. & Varmus, H. E. *Molec. cell. Biol.* **7**, 3978–3984 (1987).
10. Bradley, R. S. & Brown, A. M. *EMBO J.* **9**, 1569–1575 (1990).
11. Papkoff, J. & Schryver, B. *Molec. cell. Biol.* **10**, 2723–2730 (1990).
12. Mansour, S. L., Thomas, K. R. & Capecchi, M. R. *Nature* **336**, 348–352 (1988).
13. Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. *Nature* **309**, 255–256 (1984).
14. Capecchi, M. R. *Trends Genet.* **5**, 70–76 (1989).
15. Capecchi, M. R. *Science* **244**, 1288–1292 (1989).
16. Rossant, J. & Joyner, A. *Trends Genet.* **5**, 277–283 (1989).
17. Thomas, K. R. & Capecchi, M. R. *Cell* **51**, 503–512 (1987).

18. Doetschman, T., Maeda, N. & Smithies, O. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8583–8587 (1988).
19. Zimmer, A. & Gruss, P. *Nature* **338**, 150–153 (1989).
20. Joyner, A. L., Skarnes, W. C. & Rossant, J. *Nature* **338**, 153–156 (1989).
21. Johnson, R. S. et al. *Science* **245**, 1234–1236 (1989).
22. Koller, B. H. & Smithies, O. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8932–8935 (1989).
23. Mansour, S. L., Thomas, K. R., Deng, C. & Capecchi, M. R. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
24. Zijlstra, M., Li, E., Sajjadi, F., Subramani, S. & Jaenisch, R. *Nature* **342**, 435–438 (1989).
25. Schwartzberg, P. L., Goff, S. P. & Robertson, E. J. *Science* **246**, 799–803 (1989).
26. Zijlstra, M. et al. *Nature* **344**, 742–746 (1990).
27. DeChiara, T. M., Efstratiadis, A. & Robertson, E. J. *Nature* **345**, 78–80 (1990).
28. Koller, B. H., Marrack, P., Kappler, J. W. & Smithies, O. *Science* **248**, 1227–1230 (1990).
29. Fell, H. P., Yarnold, S., Hellstrom, K. E. & Folger, K. R. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8507–8511 (1989).
30. Schwartzberg, P. L., Robertson, E. J. & Goff, S. P. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3210–3214 (1990).
31. Hallonet, M. R., Teillet, M. & LeDouarin, N. M. *Development* **108**, 19–31 (1990).

ACKNOWLEDGEMENTS: We thank S. Tamowski, C. Lenz, M. Allen and E. Nakashima for technical assistance, T. Musci and R. Mullen for assistance and advice on histology, R. Nusse for providing the *int-1* flanking probe, D. Gard and F. Brown for help with photography.

## 'Formins': proteins deduced from the alternative transcripts of the limb deformity gene

Richard P. Woychik\*, Richard L. Maas\*, Rolf Zeller\*, Thomas F. Vogt & Philip Leder

Department of Genetics, Harvard Medical School, and Howard Hughes Medical Institute, 25 Shattuck Street, Boston, Massachusetts 02115, USA

VERTEBRATE limb formation is an evolutionarily conserved process programmed by an array of morphogenetic genes<sup>1–5</sup>. As a result of transgene insertion, we previously identified a mutation at the mouse limb deformity (*ld*) locus that disrupts embryonic pattern formation, resulting in a reduction and fusion of the distal bones and digits of all limbs as well as variable incidence of renal aplasia<sup>2,6–9</sup>. We have now characterized the *ld* locus at the molecular level. It contains evolutionarily conserved coding sequences that are transcribed in adult and embryonic tissues as a complex group of low abundance messenger RNAs created by alternative splicing and differential polyadenylation. The association of these transcripts with the gene responsible for the mutant phenotype was established by demonstrating that they are disrupted in two independently arising *ld* alleles<sup>8</sup>. We have now deduced the structure of several novel proteins (termed formins) from the long open reading frames encoded by the various *ld* transcripts. The observation of these different RNA transcripts in different tissues suggests that the formins play a part in the formation of several organ systems.

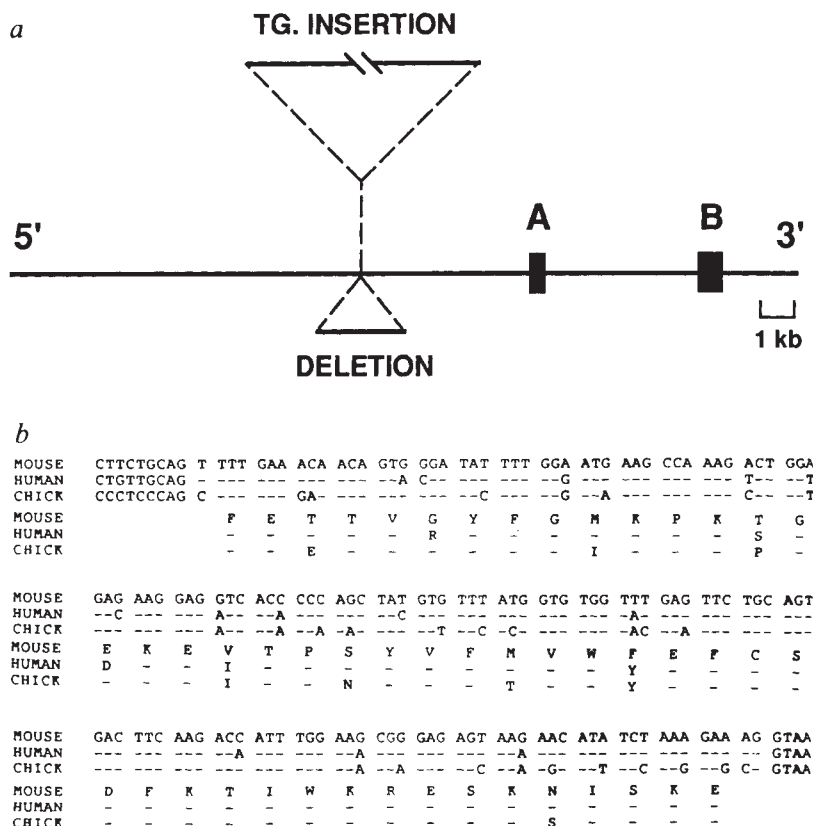
A map of both the mutant and wild-type loci is shown in Fig. 1a. Insertion of the transgene into the (*ld*) locus was accompanied by deletion of about 1.5 kilobases (kb) of endogenous

genomic DNA<sup>6</sup>. No other rearrangements were observed. We first identified potential exons along the genomic DNA by searching for evolutionarily conserved regions. Two of these appeared as likely coding exons because, when sequenced, they revealed short open reading frames flanked by splice acceptor-donor sequences (A and B in Fig. 1a). Cross-hybridizing regions were cloned from human and chicken genomic libraries and examined for sequence homology. The nucleotide sequences of regions A and B are 80–90% conserved between humans and mice. For example, the extent of sequence conservation of one putative exon (Fig. 1a, region B) is shown by comparing the homology regions of the mouse, human and chicken genomes (Fig. 1b).

Complementary DNA libraries prepared from RNA isolated from adult kidney and testis (two organs in which transcripts were most abundant) yielded overlapping cDNA clones that were identified using a genomic DNA probe containing the evolutionarily conserved exon B (shown in Fig. 1a, b). Sequence analysis of these cDNAs, illustrated by ten representative cDNA clones (Fig. 2), demonstrates a complex pattern of mRNA synthesis (see also Fig. 3). These cDNAs reveal differential use of polyA addition sites separated by several kilobases as well as alternative splicing of internal (see clones 2, 3, 4, 5, 7 and 9, Fig. 2) and 5' exons (see clone 1, Fig. 2). Long, but variably composed open reading frames (ORFs) (see below) were deduced from nucleotide sequence analysis.

The complexity of the RNA products of the *ld* locus is reflected by the various transcripts observed in analyses of RNA from both embryonic and adult tissues (Fig. 3). A transcript important in morphogenesis of limb and kidney should be present before or at the time of the morphological observation of the mutant phenotype. Expression of the *ld* gene as assayed by RNase protection occurs in primitive-streak embryos (gestational day seven) before limb development<sup>9</sup>. To examine the pattern of RNA transcripts in the embryo, RNA was isolated from gestational day 9–12 embryos and analysed as shown in Fig. 3a. Three main classes of mRNAs (~13, 7 and 3 kb) are detected in all parts of the embryo (Fig. 3a) including the limb buds. The size differences observed between these transcripts result from differential utilization of the three sets of poly(A)

\* Present addresses: Biology Division, Oak Ridge National Laboratories, PO Box 2009, Oak Ridge, Tennessee 37831, USA (R.P.W.), Department of Medicine, Brigham & Women's Hospital, Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA (R.L.M.), and EMBL, Postfach 10.2209, Meyerhofstr. 1, D-6900 Heidelberg, FRG (R.Z.).



**FIG. 1** Structure of genomic wild-type and mutant *Id<sup>Hid</sup>* DNA and comparison of evolutionarily conserved sequences in the *Id* locus. **a**, Diagrammatic representation of cloned portions of the *Id* locus. The interrupted solid line located at the upper portion of the diagram represents the inserted transgene (TG), a MMTV-cMYC fusion construct<sup>13</sup>. The deletion noted below represents ~1.5 Kb of DNA lost in the insertional *Id<sup>Hid</sup>* mutant<sup>6</sup>. Using cloned mouse genomic DNA fragments<sup>6</sup> as probes, several variably sized, evolutionarily conserved regions were detected as sequences cross-hybridizing to human genomic DNA (solid boxes A and B). These cross-hybridizing human fragments were cloned by screening a chicken genomic library prepared either in Charon 4A (ref. 14) or by ligating *Eco*RI-digested, size-selected DNA into  $\lambda$ gtWES according to standard procedures<sup>15</sup>. Cross-hybridizing fragments were cloned by screening a chicken genomic library prepared in the phage EMBL3. The homologous regions from the mouse, chicken and human genomes were sequenced using the dideoxy technique<sup>15</sup> and compared. These regions proved to be exons of the *Id* gene (solid boxes A and B, not drawn to scale, are 52 and 150 base pairs (bp), respectively). **b**, Comparison of the nucleotide sequences of exon B from the mouse, human and chicken. Nucleotide and amino acid differences of the human and chicken gene with the mouse sequence are shown; dashes represent identities. The internal 150 bp coding region is flanked by splice donor and acceptor dinucleotides<sup>16</sup>.

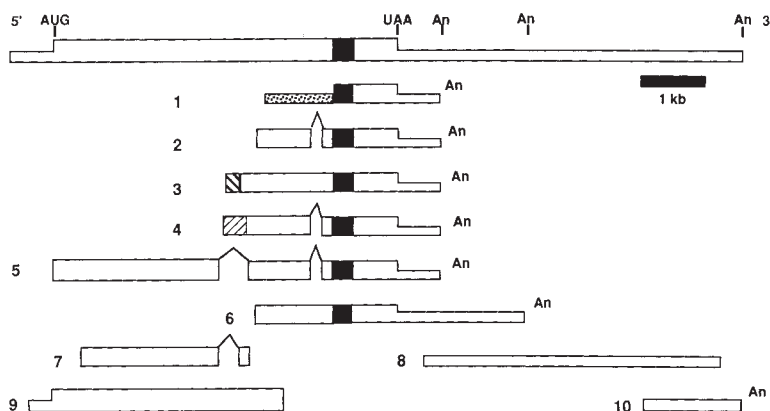
addition sites and alternative splicing of exons (Fig. 2). The *Id* transcripts are relatively rare; levels in embryos are roughly 200-fold less than actin mRNAs<sup>9</sup>.

Many adult organs also show a complex pattern of *Id* RNA expression. For example, the testis contains at least six different sizes of transcripts (Fig. 3b). The expression pattern of the *Id* gene in ovary, kidney, brain, small intestine, salivary gland and Harderian gland is of similar complexity. Much simpler expression patterns (or no expression at all) are seen in other

adult organs analysed (Fig. 3b). In addition, RNase protection analyses reveal that alternative splicing occurs (not shown). We have found, for example, distinct splicing patterns: one common to brain and testes, another common to small intestine and salivary gland. Both patterns are present in the kidney. Although variation of splicing is seen among different tissues, no clear lineage restriction (for example to ectoderm, mesoderm or endoderm) of *Id* gene expression is evident.

To gain some notion of the protein products encoded by the

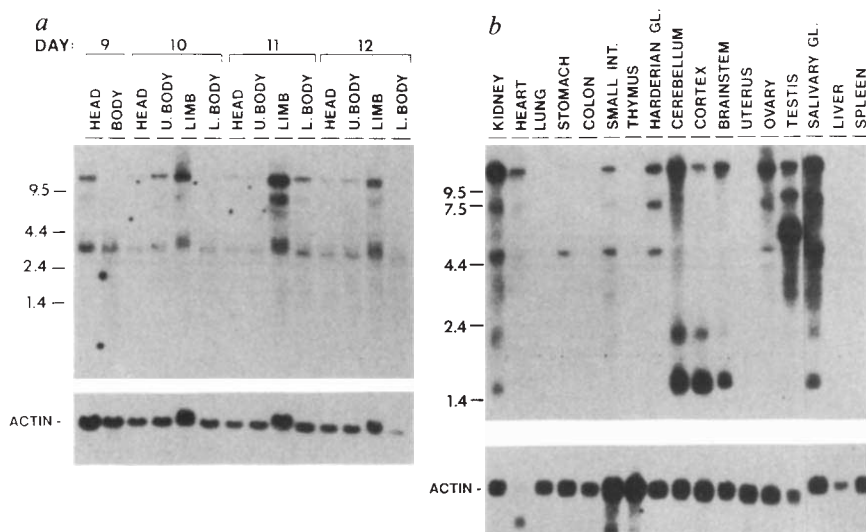
**FIG. 2** Structure of the *Id* mRNA family as deduced from representative cDNA clones. A composite mRNA structure (top) covering ~13 kb is based on the characterization of 10 different overlapping cDNA clones isolated from adult testis and kidney mRNA (numbers to the left). The large size of the transcripts precluded the isolation of full-length clones from any transcript. Three different main sites of polyadenylation (An), separated by several kb, as well as many alternative forms of splicing were apparent on the cDNAs. A 1,468-amino-acid ORF, shown as an open region extending from the AUG to UAA, was assembled from the overlapping ORF's in clones 3, 6 and 9. The thin regions flanking the AUG and UAA represent the predicted 5' and 3' noncoding portions of the transcripts. Utilization of a different acceptor site to generate an alternative splice in the middle of clone 5 shifts the reading frame in this clone and introduces a termination codon in the third triplet following the splice. Filled solid regions, juxtaposed exons A and B shown in Fig. 1. In clone 1 the stippled region represents unique sequences transcribed from the 3' side of the transgene insertion site and appears only in RNA derived from testis. Clone 3 bears a 170 nucleotide (nt) 5' sequence (hatched box), present as low abundance RNA and only in the testis. The hatched region at the 5' end of clone 4 represents sequences unique to this clone. To rule out cloning artefacts, the authenticity of these cDNAs has been verified using total cellular RNA in RNase protection assays (data not shown). To illustrate subtle structural details, portions of the diagram are not drawn to scale. **METHODS.**  $\lambda$ gt10 cDNA libraries were prepared with poly(A)<sup>+</sup> RNA from either CD1 adult mouse testicular or kidney tissue using standard pro-



cedures<sup>15</sup>. Initially, 400,000 plaques were screened using a small genomic fragment containing exon B (Fig. 1) as a probe. This yielded seven positive clones, six of which were characterized. Additional rounds of screening using several million plaques from both the kidney and testis cDNA libraries were used to extend the overlapping cDNA sequence in both the 5' and 3' directions. The 10 cDNA clones represent only a portion of the total clones that were isolated. Clones 1, 2, 3 and 7 were sequenced completely using the dideoxy method<sup>15</sup>. Several regions of the other clones were sequenced by extension from synthetic oligodeoxynucleotide primers to derive the indicated structures.



FIG. 3 *ld* mRNAs analysed by size fractionation and blot hybridization. *a*, Poly(A)<sup>+</sup> RNA (~7 µg) isolated from the limb, head, body of the day-9, -10, -11 or -12 developing mouse fetus, or *b*, poly(A)<sup>+</sup> RNA (~3 µg) isolated from various indicated tissues of the adult, were electrophoresed and blotted as described<sup>1,5</sup> and probed with a 528 bp cDNA fragment corresponding to the 5' terminus of clone 3 (Fig. 2). Sizes of RNA markers are indicated (kb). Filters were rescreened for actin mRNA to assure integrity of the RNA sample<sup>1,7</sup> (see bottom panel).



```

1  MEGTHCTLQL HNPFAELCYI SFYLPKGEVR GFSYKGTVTL DRSNNAFHNC
51  YQVREGPDIT SLSQPNNEHP GDIFPQFTPT KNILTELYKL TAEKERLLDS
101 LLRSNDILGV SMGSQEQKLQ ELSVILATGD EYFQSAQNR RELPVSSILIR
151 RSTQENKKPR RSGRRRESPE ELRQKTRRRK GRGQESAFQ MGKQVQCSSS
201 SLSFRANPNL RLLEERGNLA PRGTLTSSLR RRESCPANIL RPDADLAFLG
251 NSGRTEEDTD LEGPLSPDSS PTEVGADAVG GQKSSSHQOE PPQPNVSSEH
301 GKHAERWRS SRTKSKSLE RTCSKKPVSK VVAKIQEPSA PVKRIVRAHH
351 DGKGRVAYGP ETQTEFIPKA DFLTLPGGET ETHSSGRLEB EQPGIKSLRS
401 SAPERASITK EPASTEAAVN KVLRLKVESE KLDEATEGRF LGFSLNTRAT
451 HTFPETRSQR KAGLPQSGHK FLLDLPLHTV GPDSPPQKCD EKKPTQVPT
501 ALGMVFNNSS PQSSAHKRLS PVPSPLSPRC PSPQQHRLIL LLPPLSEGE
551 VVFNEYPSRK NDVSSGFPSA DTLEPSSTK VTETKGASPT SLRASQTWLV
601 SEEASEKGLG PEKITAPPQH QLPPGIASEG FPCDNFKEQT AKDLNPKDGG
651 VVWPGYRAGP PCPFLHEEK EKTSSRELYL DLNPDQSPTE QDRTPGRLQ
701 AVWPPPKTKD TEEKVGLKYT EAEYQAAILH LKREHKEELE TLQAQFELKT
751 FHIRGEHALV TARLEEAIEI LKQLEKRRR GCEEMRDVCI STDDCSPKA
801 FRNVCQTDR ETFLKPCDAE SKATRSSQIV PKLTIISLTQ LSPSDSKDI
851 HAPFQIREGT SSSSQKISP PAPPTPPPLP PPLIPPPPLP PPGGLPLPPA
901 PPIPPVCPVS PPPPPPPPPP TPVPPSDGPP PPPPPPPPLP NVLALPNSGG
951 PPPPPPPPPP PPGGLAPPPP GLSGLSSSS SQYPRKPAIE PSCPMKPLYW
1001 TRIQINDXSQ DAAPTLDWDL EEPHIRTSE FEYLFSDTT QKKKKPLSEA
1051 YEKNKVKKI IKLLDGKRSQ TVGILISSHL LEMKDIQAI FTVDSDVVDL
1101 ETLAALYENR AQEDLTQIR KYVETSKEDD LKLLDKPEOF LHLELAIPNF
1151 AERAQCIIFR AVFSEGITSL HRKVEIVTRA SKGLLHMKSIV KDILALILAF
1201 GNYMNGNRT RGQADGYSLE ILPKLDKVS RONGMNLVDY VVKYLYRYD
1251 QCKHHDQEAS CRGKDLFSLY FHIAVHPQK SOLELRQEAG TDKSVFLPE
1301 PQDFFLASQV KFEDLLKDLR KLRQLEASE QMKLVCKES PREYLQPFKD
1351 KLEEFKKAH KEHMEESHL ENAQSFETT VGYFGMKPPT GEKEVTPSYV
1401 FMVWFECSD FKTINKRESK NISKERLKMA QASVSKLTSE KKVETKINP
1451 TASLKERLRQ KEASVATN*

```

FIG. 4 Primary sequence of a putative 164K protein deduced from the 1,468-amino-acid ORF derived from the overlapping cDNA clones 3, 6, 7 and 9 (Fig. 2). Portions of the ORF boxed within dashed lines (amino acid positions 625-722 and 723-743) represent sequences present in cDNA clones 3 and 9, but excluded by alternative splicing in clones 4, 5 and 7 (see Fig. 2). The ORF derived from cDNA clone 5 encodes a 69K protein whose sequence is identical to that shown up to amino-acid position 625, followed by two additional amino-acid residues, ser-val. Clone 4 (Fig. 2) encodes an alternative amino-acid sequence, N-terminal to position 684 (not shown). The boxed region within dashed lines from amino-acid positions 1,251-1,286 represents an alternative sequence present in clones 3 and 6, but absent in clones 2, 4 and 5. A proline-rich region is evident from positions 870-970. A sequence which would be lost by truncation as a consequence of the transgene insertion in the *ld*<sup>10</sup> allele is boxed within solid lines (positions 1,360-1,468). The two exons, A and B, shown in Fig. 1 are from amino-acid positions 1,360-1,376 and 1,377-1,425, respectively.

*ld* gene, a composite mRNA sequence was derived from several of the cDNA clones shown in Fig. 2. The composite structure derived from overlapping cDNA clones 3, 6, 7 and 9 yields an ORF that can be translated into a 1468 amino-acid protein with a predicted relative molecular mass of 164,000 ( $M_r$ , 164K) (Fig. 4). An ATG codon near the 5' end of clone 9 was chosen as the likely translational start site on the basis of the presence of stop codons in all three reading frames 5' to this position and the occurrence of a sequence that matches the initiation consensus sequences proposed by Kozak<sup>10</sup> at all important positions. Other protein products can be postulated by using alternative splicing. For example, the splicing pattern reflected in cDNA clone 5 (Fig. 2) would lead to a change in the ORF resulting in a protein of 69K that is truncated two codons after the first splice junction. Consistent with this, RNA transcribed from composite cDNA clones can be translated into appropriately sized polypeptides in an *in vitro* translation system (T.F.V., unpublished data).

A comparative search of protein sequence databases failed to identify any significant homologies to other known proteins or amino acid motifs. The deduced 164K protein lacks any obvious signal peptide or transmembrane segment, but encodes a proline-rich core (positions 870-970) that might serve as a molecular hinge, dividing the protein into two main domains. The N-terminal domain is relatively abundant in basic amino acids, and the C-terminal domain is rich in potential  $\alpha$ -helical structure. Two zinc-finger-containing, putative DNA binding transcription factors, Krox20 and the predicted product of a candidate Wilms' tumour gene, contain proline stretches<sup>11,12</sup>. The composite ORF shown in Fig. 4 represents only one of several potential gene products created by alternative splicing of *ld* transcripts (see Figs 2 and 4, where alternative exons are boxed within dashed lines). As these apparently novel proteins seem essential for the proper formation of both limbs and kidneys<sup>7-9</sup>, we have named them the formins. An obvious function does not yet emerge from the structure of the *ld* ORF we have determined. Further analysis involving tissue distribution and cellular localization may allow us to place the formin proteins in a functional category. In any case, our molecular characterization of the *ld* locus provides a basis to begin functional analysis of the formins and their alterations in the various mutant *ld* alleles<sup>8</sup>.

Received 13 April; accepted 14 June 1990.

- Hinchliffe, J. R. & Johnson, B. R. *Development of the Vertebrate Limb* (Clarendon, Oxford, 1980).
- Green, M. C. *Genetic Variants and Strains of the Laboratory Mouse* (eds Lyon, M. F. & Searle, A. G.) 200-201 (Gustav Fischer Verlag, New York, 1989).
- Gruneberg, H. *The Pathology of Development* (Wiley, New York, 1963).
- Saunders, J. W. Jr *J. exp. Zool.* **108**, 363-404 (1948).

5. Tickle, C., Summerbell, D. & Wolpert, L. *Nature* **254**, 199–202 (1975).
6. Woychik, R. P. *et al.* *Nature* **318**, 36–40 (1985).
7. Kleinbrecht, J., Selow, J. & Winkler, W. *Anat. Anz.* **152**, 313–324 (1982).
8. Maas, R. L. *et al.* *Nature* **346**, 853–855 (1990).
9. Zeller, R., Jackson-Grusby, L. & Leder, P. *Genes Dev.* **3**, 1481–1492 (1989).
10. Kozak, M. *Nucleic Acids Res.* **12**, 857–872 (1984).
11. Chavrier, P. *et al.* *EMBO J.* **7**, 29–35 (1988).
12. Gessler, M. *et al.* *Nature* **343**, 774–778 (1990).
13. Stewart, T. A., Pattengale, P. K. & Leder, P. *Cell* **38**, 627–637 (1984).
14. Maniatis, T. *et al.* *Cell* **15**, 687–701, 1978.
15. Ausubel, F. M. *et al.* *Current Protocols in Molecular Biology* (Wiley, New York, 1988).
16. Breathnach, Z. *et al.* *Proc. natn. acad. Sci. U.S.A.* **75**, 4853–4857 (1978).
17. Minty, A. M. *et al.* *J. biol. Chem.* **256**, 1008–1018 (1981).

**ACKNOWLEDGEMENTS.** We thank K. van Myrik and M. Fishman for their participation in the early aspects of these studies. We also thank L. Jackson-Grusby for helpful comments on this manuscript. The chicken genomic library was a gift from D. Engel. This work was supported in part by the E.I. DuPont de Nemours and Co., Inc. R.P.W. was supported by the Jane Coffin Childs Memorial Fund for Medical Research, R.Z. was supported by an EMBO Fellowship and the Swiss National Science Foundation and T.F.V. was supported by the American Cancer Society. Research jointly sponsored by the Office of Health and Environmental Research, US Department of Energy with Martin Marietta Energy Systems, Inc. and by additional funding sources as listed above. The composite nucleotide sequence has been deposited in the EMBL/GenBank database (Bolt, Beranek and Newman Laboratories, Cambridge, MA and European Molecular Biology Laboratories, Heidelberg (Accession Number X53599, mouse limb deformity cDNA)).

## Disruption of formin-encoding transcripts in two mutant limb deformity alleles

Richard L. Maas\*, Rolf Zeller\*, Richard P. Woychik\*,  
Thomas F. Vogt & Philip Leder

Department of Genetics, Harvard Medical School,  
and Howard Hughes Medical Institute, 25 Shattuck Street, Boston,  
Massachusetts 02115, USA.

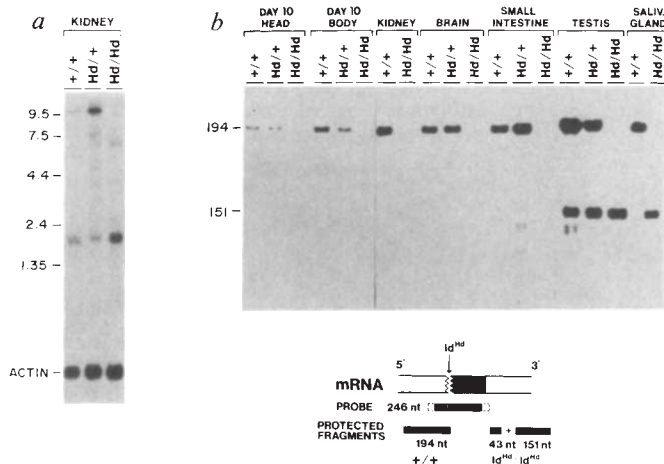
**THE** recent identification of a gene residing at the mouse limb deformity (*ld*) locus<sup>1–5</sup> permits us to test the hypothesis that disruption of this gene is responsible for an inherited anomaly affecting embryonic pattern formation. The gene gives rise to alternatively processed messenger RNAs that can be translated as a family of related protein products, termed the formins<sup>1</sup>. We have now analysed transcripts from this gene in four independently isolated mutant alleles. In two of these, the *ld*<sup>Hd</sup> allele (created by insertion of a transgene<sup>2</sup>) and the *ld*<sup>In2</sup> allele (created by a translocation–inversion involving mouse chromosomes 2 and 17, ref. 6), a common subset of *ld* transcripts is abolished, but others are apparently unaltered. The correlation of altered transcripts in two independent *ld* mutants strongly supports the notion that one or more altered formins is responsible for the observed phenotype. That the defect is limited to the limb and kidney, despite expression of *ld* mRNA in other unaffected organs, suggests that these mutant alleles represent only partial loss of *ld* function.

Homozygous transgene-bearing (*ld*<sup>Hd</sup>) mice lack some, but not all, *ld* mRNAs that are present in heterozygous and wild-type animals. For example, the predominant 13-kilobase (kb) transcripts are missing from the kidneys of homozygous mice (Fig. 1a), but the 2 kb transcripts seem unaffected. To analyse more precisely the disruption of *ld* mRNAs in transgenic mice, a complementary probe incorporating exons from both sides of the transgene insertion site was used in RNase protection analyses (see map in Fig. 1b). This probe detects two normally joined exons between which an intronic insertion of the transgene has occurred in the *ld*<sup>Hd</sup> mutation. Messenger RNA molecules bearing these joined exons should protect a 194-nucleotide (nt) sequence. As shown in Fig. 1b, RNAs derived from tissues of adult and embryonic wild-type and heterozygous mice protect

the correctly spliced sequence. If the transgene insertion interferes either in the transcription or splicing of these exons, RNA derived from homozygous mutant mice should not yield this protected fragment. That this is the case is shown in Fig. 1b. Moreover, this analysis also shows a complete absence of correctly spliced wild-type RNA in day 10 mutant mouse embryos (the stage at which the *ld* phenotype first appears<sup>4</sup>). Similar results were obtained using RNA from stages of mouse embryogenesis through day 13 (data not shown).

An additional mutant *ld* allele, *Is(17; In2)ld*, *a*<sup>1</sup>Gso or *ld*<sup>In2</sup>, showing both a 2; 17 chromosomal translocation and a structural rearrangement of chromosome 2, has been identified<sup>6</sup>. Analysis of *ld* RNA transcripts from these mutant mice shows that they are also altered (Fig. 2a). For example, a cDNA probe detects a 13-kb class of RNAs in wild-type brain, kidney, testis and salivary gland, which is absent in homozygous mutant *ld*<sup>In2</sup> tissues. In addition, a 7-kb RNA normally seen in testis and salivary gland is missing or severely diminished in mutant organs.

RNase protection analysis of RNA from heterozygous (+/*ld*<sup>In2</sup>) and homozygous (*ld*<sup>In2</sup>/*ld*<sup>In2</sup>) mice is also consistent with the differences noted above (Fig. 2b). For example, a 477-nt RNA sequence incorporating correctly spliced exons from either side of the main structural rearrangement (see map, Fig. 2b) is clearly absent in homozygous animals. However, the portion 5' to the translocation breakpoint (reflected by a protected 256-nt



**FIG. 1** *a*, Analysis of poly(A)<sup>+</sup> RNA from kidney of wild-type, heterozygous *ld*<sup>Hd</sup>/+, and homozygous *ld*<sup>Hd</sup>/*ld*<sup>Hd</sup> mice. Numbers refer to size markers (kb). The probe was a nick-translated 1.3-kb *EcoRI*–*HindIII* cDNA fragment; similar results were obtained with two other cDNA probes. To control for amount and integrity of the RNA the filter was stripped and rehybridized with a probe for actin mRNA as shown in the bottom panels<sup>13</sup>. Poly(A)<sup>+</sup> RNA (2 µg) isolated from various tissues was prepared, electrophoresed and blotted as described<sup>14</sup>. Autoradiography was for 3–7 days at –70 °C with one intensifying screen. *b*, RNase protection analyses of RNA from wild-type, heterozygous and homozygous *ld*<sup>Hd</sup> adult and embryonic mice. Embryos were staged by counting the day of plug as day 0. Genotypes of embryos were determined using the dot blot assay described previously<sup>4</sup>. The antisense probe was a 194-bp *PstI*–*PstI* cDNA fragment subcloned into the *PstI* site of pGEM3 (Promega), linearized with *EcoRI* and transcribed using SP6 polymerase, resulting in a probe of 246 nt. The solid region of the probe represents cDNA sequence and the open flanking region represents vector sequences. Fragment sizes in nucleotides are shown. Embryonic day-10 (20 µg) or adult tissue (40 µg) total RNA was hybridized to the probe and analysed by a standard RNase protection assay as described<sup>15</sup>. A 151-nt fragment is protected by RNA derived from the testes of wild-type, heterozygous and homozygous mice and from the salivary gland of homozygous animals. In RNA from mutant testis and salivary glands these exons are transcribed principally as the result of tissue-specific expression of the MMTV transgene (data not shown). In the case of wild-type testis, the 151-nt fragment is protected by a transcript presumed to be initiated 3' to the insertion site (see Fig. 2, clone 1, in Woychik *et al.*, 1990).

\* Present Addresses: Department of Medicine, Brigham & Women's Hospital and Harvard Medical School, Howard Hughes Medical Institute, 75, Francis Street, Boston, Massachusetts 02115, USA (R.L.M.), EMBL, Meyerhofstr. 1, D-6900 Heidelberg, FRG (R.Z.) and Biology Division, Oak Ridge National Laboratory, PO Box 2009, Oak Ridge, Tennessee 37831, USA (R.P.W.).



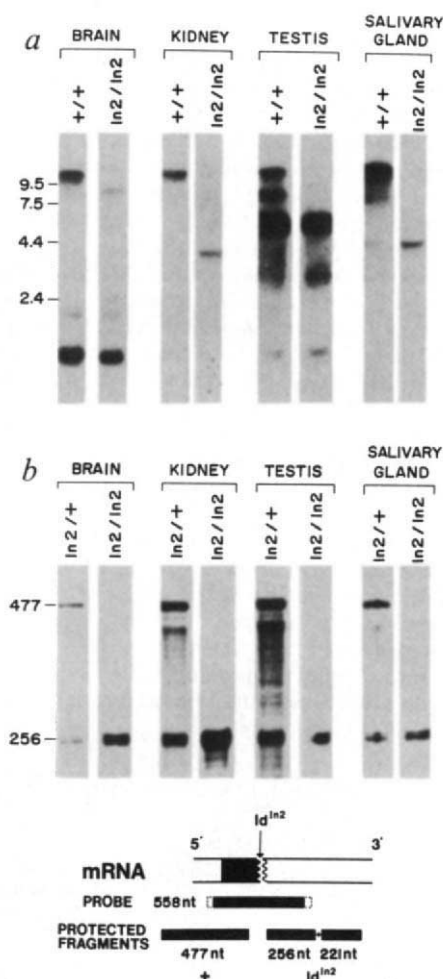


FIG. 2 Analyses of RNA isolated from different organs of wild-type and mutant  $ld^{ln2}/ld^{ln2}$  adult mice. *a*, Northern blot analysis. Samples contained poly(A)<sup>+</sup> RNA (~5  $\mu$ g) isolated from the indicated organ from either wild-type (+/+) or homozygous ( $ld^{ln2}/ld^{ln2}$ , labelled ln2/ln2 in the Figure) mice, using the methods described<sup>14</sup>. The cDNA probe used to detect RNA is a 528-bp *Pst*I–*Eco*RI cDNA fragment isolated from the mid-portion of the coding region. Numbers refer to size markers (kb). *b*, Ribonuclease protection analysis. The antisense probe was a 477-nt *Sty*I–*Spe*I fragment subcloned into pBluescript II KS(+) (Stratagene), linearized with *Hind*III and transcribed with T7 polymerase, resulting in a 558-nt probe. The solid region of the probe represents cDNA sequence and the open flanking regions represent vector sequences. The antisense RNA probe used is shown, with the expected protected fragment sizes in nucleotides for wild type (+) and mutant ( $ld^{ln2}$ ) alleles. The analysis was carried out as described<sup>15</sup> using total RNA (30  $\mu$ g) from either heterozygous (ln2/+) or homozygous (ln2/ln2) tissues.

fragment) is transcribed in homozygous mutant animals at a level comparable to that in wild-type. The absence of the 3' portion of this sequence (as reflected by the absence of a 221-nt protected fragment) suggests that sequences 3' to the rearrangement are either not transcribed or are highly unstable.

The finding of absent and altered transcripts in two independent *ld* mutants provides strong evidence that mutations in this gene are responsible for the mutant phenotype. Two additional points are consistent with this conclusion. First, the *ld* transcripts arise from the only transcription unit that we can identify using RNase protection assays and evolutionarily conserved homology regions as probes<sup>1</sup>. Second, the expression of this gene fulfills one of the central requirements necessary for its involvement in the phenotype: it is expressed in the developing embryo at or before the stage (day 10) at which phenotypic abnormalities can first be detected<sup>4</sup>.

A further question arises as to how the relatively broad transcription expression pattern of the *ld* gene is reconciled with a phenotype that seems restricted to the limb and kidney. One model that might account for this observation would be that the broadly transcribed *ld* gene products only function in a subset of the expressing tissues. Examples of loci that are expressed in a broader range of cell types than those affected by null mutations are the *Drosophila* genes; *sevenless*<sup>7,8</sup>, *Notch*<sup>9</sup> and *pole-hole*<sup>10</sup>. Alternatively, *ld* function might be compensated for in unaffected tissues by functional redundancy provided by other loci. In the nematode, for instance, certain collagen mutants are likely to be functionally compensated by other collagen types<sup>11,12</sup>. Our characterization of the *ld* gene raises a third possibility; namely, that the observed phenotypes are produced by a partial loss of function or by the loss of one of several

relatively independent functions. The different transcripts might encode functionally different formins, and thus mutations at different positions (for example, in exons incorporated in a tissue-specific fashion) might affect function in only a subset of formin products, leaving others unaffected. We do not know whether any of the *ld* alleles represent the null phenotype (that is, complete loss of function).

Although we have identified two independent *ld* mutant alleles that share a common subset of absent transcripts, we have not detected alterations in *ld* mRNA expression in either of the other mutant alleles ( $ld^1$  or  $ld^{OR}$ ) (ref. 3). As these alleles have only been examined at the level of restriction mapping of genomic DNA, northern-blot and RNase-protection analyses of adult tissue RNA, and limited nucleotide sequence comparisons (data not shown), a small alteration in sequence would go undetected. Identification of the mutations in these alleles producing essentially the same phenotype may reveal a more delimited mutational event (such as a single base change), and thus allow us to pinpoint a specific version of the alternative forms of *ld* mRNAs as that responsible for the observed phenotype. Advances in the manipulation of mouse embryonic stem cells present the opportunity to carry out site-directed mutagenesis of the *ld* gene, thus allowing the construction of mutants to test this possibility. □

Received 13 April; accepted 14 June 1990.

1. Woychik, R. P. et al. *Nature* **346**, 850–853 (1990).
2. Woychik, R. P. et al. *Nature* **318**, 36–40 (1985).
3. Green, M. C. *Genetic Variants and Strains of the Laboratory Mouse*, (eds Lyon, M. F. & Searle, A. G.) 200–201 (Gustav Fischer Verlag, New York, 1989).
4. Zeller, R., Jackson-Grusby, L. & Leder, P. *Genes Dev.* **3**, 1481–1492 (1989).

5. Kleinebrecht, J., Selow, J. & Winkler, W. *Anat. Anz. Jena* **152**, 313–321 (1982).
6. Woychik, R. P. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2588–2592 (1990).
7. Tomlinson, A. et al. *Cell* **51**, 143–150 (1987).
8. Banerjee, U. et al. *Cell* **51**, 151–158 (1987).
9. Hartley, D., Xu, T. & Artavanis-Tsakonas, S. *EMBO J.* **6**, 3407–3417 (1987).
10. Ambrosio, L., Mahowald, A. P. & Perrimon, N. *Nature* **342**, 288–291 (1989).
11. Kusch, M. & Edgar, R. S. *Genetics* **113**, 621–639 (1986).
12. Kramer, J. M. et al. *Cell* **55**, 555–565 (1988).
13. Minty, A. M. et al. *J. biol. Chem.* **256**, 1008–1014 (1981).
14. Ausubel, F. M. et al. *Current Protocols in Molecular Biology* (Wiley, New York, 1988).
15. Kreig, P. & Melton, D. *Meth. Enzymol.* **155**, 397–415 (1987).

ACKNOWLEDGEMENTS. We thank L. Jackson-Grusby for helpful comments on this manuscript. This work was supported by a grant from the E.I. DuPont de Nemours & Co., Inc. and by the Office of Health and Environmental Research, US Department of Energy with Martin Marietta Energy Systems, Inc. In addition, R.P.W. was supported by the Jane Coffin Childs Memorial Fund for Medical Research, R.Z. by EMBO and the Swiss National Science Foundation and T.F.V., by the American Cancer Society.

## Peripheral organs control central neurogenesis in the leech

Carlos A. Baptista\*, Timothy R. Gershon  
& Eduardo R. Macagno†

Department of Biological Sciences, Columbia University, New York, New York 10027, USA

INTERACTIONS between developing nerve centres and peripheral targets are known to affect neuronal survival and thus regulate the adult number of neurons in many systems<sup>1,2</sup>. Here we provide evidence that peripheral tissues can also influence cell numbers by stimulating the production of neurons. In the leech *Hirudo medicinalis*, there is a population of several hundred neurons that is found only in the two segmental ganglia that innervate the genitalia<sup>3</sup> and which seem to be added gradually during post-embryonic maturation<sup>4,5</sup>. By monitoring 5-bromo-2'-deoxyuridine incorporation immunohistochemically, we have now determined that these neurons are actually born late in embryogenesis, well after all other central neurons are born and after efferent and afferent projections are established between these ganglia and the periphery. Ablation of the male genitalia early in embryogenesis, or evulsion of the nerves that connect them to the ganglia, prevent the birth of these neurons. However, they fail to appear ectopically when male genitalia are transplanted to other segments, despite innervation by local ganglia. We conclude that the generation of the late-appearing neurons depends on a highly localized signal produced by the male genitalia, to which only the ganglia that normally innervate these organs have the capacity to respond.

Leeches are segmented hermaphrodites, with the male and female reproductive organs present respectively in the fifth and sixth body segments. Their central nervous system (CNS) comprises 21 segmental ganglia (SG1–SG21), as well as head and tail ganglia. The segmental ganglia that innervate the sexual organs are SG5 and SG6 and are also designated the sex SG. Both ganglia innervate the male organs; only SG6 innervates the female organs (Fig. 2a). Whereas most SG in adult *Hirudo* contain about 400 neurons, the sex SG contain an additional population of ~350 small neurons<sup>3</sup>. These are distributed throughout each sex ganglion and can be identified by their small size and, although not exclusive to these neurons, by their expression of Phe-Met-Arg-Phe-NH<sub>2</sub> (FMRFamide) immunoreactivity<sup>5,6</sup>. As the number of extra neurons increases gradually during the first year of postembryonic life<sup>4,5</sup>, they were initially called PE (postembryonic) neurons. But, as reported below, we have now determined that these neurons are actually born during late embryogenesis, and that their genesis is control-

TABLE 1 Numbers of BrdU-labelled cells in animals with transplanted male genitalia

|         | n  | SG4     | SG5       | SG6       | SG7      |
|---------|----|---------|-----------|-----------|----------|
| Normal  | 14 | 9.9±1.8 | 50.2±8.7  | 46.9±6.3  | 5.7±2.3  |
| MGT-SG6 | 4  | 8.0±2.2 | 8.3±1.7   | 74.5±10.4 | 7.3±2.5  |
| MGT-SG7 | 3  | 9.0±1.9 | 64.7±29.0 | 66.0±21.4 | 11.0±2.6 |

Transplants were at E10–E11, the exposure to BrdU at E24, and the immunostaining one day after the exposure. The numbers indicate the mean and s.e.m.; n is the number of specimens analysed. In MGT-SG6 animals (male genitalia transplanted close to SG6), the host's male and female genitalia were removed before the transplantation of donor tissue; hence, these animals only had the transplanted male genitalia, and these were innervated only by SG6. In MGT-SG7 animals (male genitalia transplanted close to SG7), the host genitalia were left intact; these animals had two sets of male genitalia, with the transplanted tissue innervated by SG7.

led by a peripheral target organ; hence, we have termed them peripherally induced central (PIC) neurons.

Cell counts in juveniles and adults have shown that PIC neurons are absent when the male genitalia are ablated or disconnected from the sex SG before day 16 of embryogenesis (E16) (*Hirudo* embryogenesis takes ~30 days at 22–23 °C). Transplantation of the male genitalia to body segments other than 5 and 6 at E10 does not result in the appearance of extra neurons in those segments, even though the local SG do form functional projections with the transplanted tissues<sup>5,7</sup>. Hence, the capacity to generate PIC neurons seems to be limited to the sex SG by E10. These observations could be explained if the male genitalia controlled either the genesis of the extra neurons, their survival, or their migration into the sex SG from outside the CNS. To distinguish among these possibilities, we examined the distribution of mitotic cells in the CNS of normal leeches and those whose male genital primordia had been ablated or disconnected from the sex ganglia at E10.

Mitotic activity was monitored by detecting the nuclear incorporation of 5-bromo-2'-deoxyuridine (BrdU) immunohistochemically in whole-mount preparations<sup>8–10</sup>. In these experiments, BrdU was administered either *in vivo*, by injection into the circulatory system of intact animals (E24 or older—it was not possible to inject into the circulatory system of younger embryos), or *in vitro*, by addition to the medium in which dissected embryos (E15–E25) or postembryonic nerve cords were cultured for short periods of time. The number of labelled cells was assessed at times ranging from an hour to several months after the exposure to BrdU. As control animals injected with BrdU and allowed to survive to adulthood exhibited normal morphology and apparently normal behaviour, we presume that exposure to the BrdU concentrations used in these experiments had no deleterious effects.

Dividing cells were found in significantly larger numbers in the sex SG than in adjacent ganglia of embryos exposed to BrdU at E20 for 1 day in culture (SG5, 6: 20.3±1.4 labelled cells; SG4, 7: 8.3±0.8 labelled cells; mean±s.e.m., n=21) or injected at E24–E26 (Figs 1a, b and 2b). In contrast, embryos exposed in culture at E15 (SG5, 6: 12.8±1.0 labelled cells, n=17; SG4, 7: 10.4±1.0 labelled cells, n=18; difference not statistically significant) or injected at E30 and postembryonic day 10 (P10) (Fig. 2b), showed roughly equal numbers of labelled cells in all the ganglia examined. No BrdU-immunoreactive cells were observed within ganglia of eight animals injected with BrdU at P35, nor in those of one ~3-month-old juvenile that was cultured in the presence of BrdU for one day. Thus, E20 to E26 were the only ages in which greater mitotic activity was detected in the sex ganglia than in other segmental ganglia.

To demonstrate that the additional BrdU-immunoreactive cells in the sex SG become the PIC neurons, E23–25 embryos were injected with BrdU and allowed to survive for several months, until an age when the PIC neurons express detectable

\* Present address: Department of Pathology, College of Physicians & Surgeons, Columbia University, New York, New York 10032, USA.

† To whom correspondence should be addressed.



levels of FMRFamide immunoreactivity<sup>5</sup>. There were many instances of small neurons labelled by both anti-BrdU and anti-FMRFamide antisera in the sex SG of these animals (Fig. 1e, f), indicating that it is indeed the PIC neurons that are born late in embryogenesis.

The hypothesis that the male genitalia control the birth of the PIC neurons was tested directly by examining BrdU incorporation in a group of late embryos that had had their male genitalia ablated at E10. A second group had only the nerves connecting SG6 to the male genitalia evulsed at this time, and a control group had the female genital primordia removed. All of these animals were injected with BrdU at E24–25 and immunostained a day later. The additional BrdU-immunoreactive cells were clearly absent from both sex SG in the group lacking male genitalia and from only SG6 in the group with SG6 disconnected, whereas the control group showed the normal segmental differences in labelled cell numbers (Figs 1c and 2c).

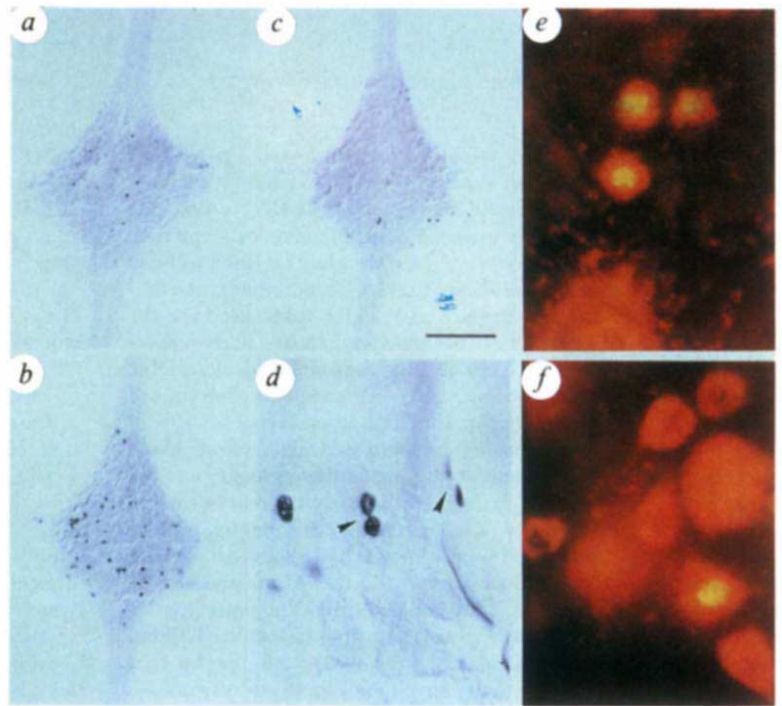
We also transplanted male genital primordia to other segments

in E10 embryos, to test the capacity of these tissues to induce cell division ectopically. Despite innervation of the transplant by the local SG, we saw no increase in the number of BrdU-immunoreactive cells (Table 1). In contrast, the ablation of the normal male genitalia and the replacement of the female by male genital primordia in body segment 6 at E10 did result in additional cells in SG6 (Table 1), showing that the surgical procedures did not affect the competence of the transplant for stimulating cell division. By E10, therefore, only the sex SG seem to have the potential for generating the PIC neurons.

We frequently observed isolated pairs of labelled cells within the sex SG at short time intervals (2 h or less) after BrdU exposure (Fig. 1d), suggesting that these cells must have been born locally, instead of being born elsewhere and migrating into the ganglia. The results of another set of experiments also supported this conclusion; in those experiments, four E24 animals were exposed to BrdU for 12 h in culture after the nerves of one of the two sex SG and one adjacent non-sex ganglion

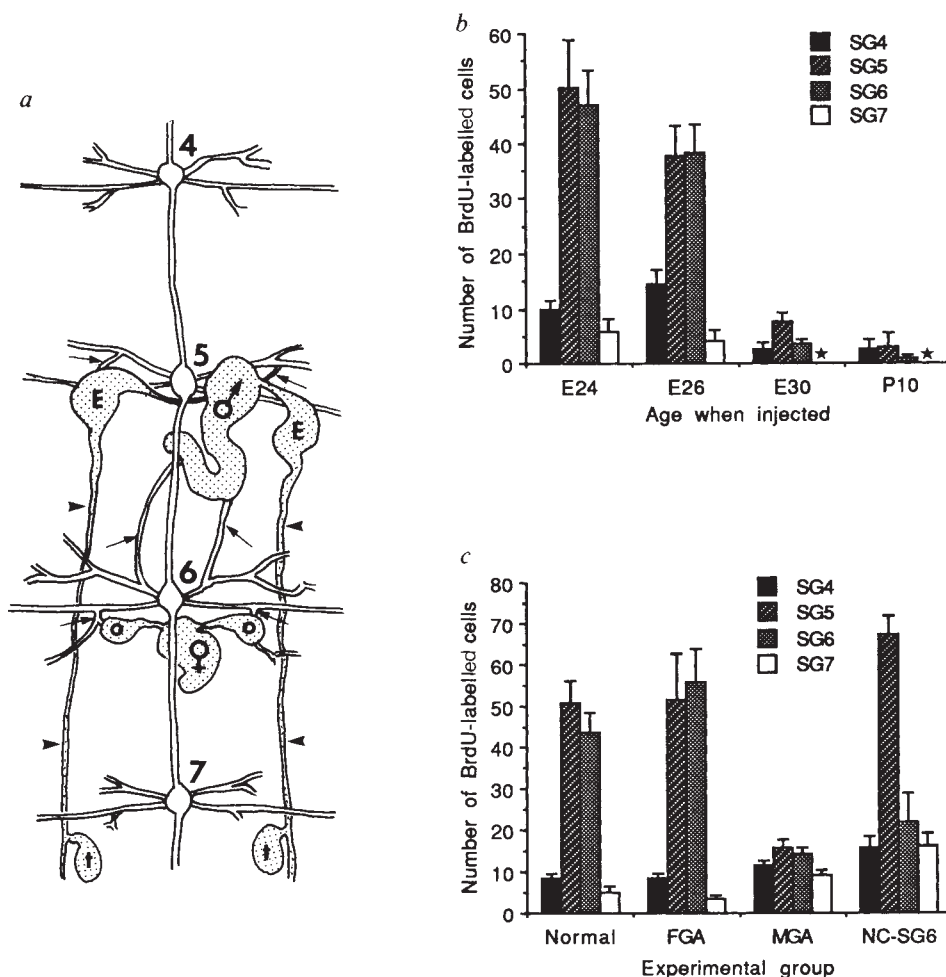
FIG. 1 Segmental differences in neurogenesis during late embryonic development generate the PIC neurons within the sex ganglia. Micrographs of ganglia from whole mounts stained with immunoperoxidase show incorporation of BrdU (a–d), and stained with immunofluorescence show simultaneous incorporation of BrdU and of FMRFamide immunoreactivity (e, f). All animals were injected with BrdUTP at E24. a, Representative non-sex ganglion (SG4) from a normal E25 embryo. b, Sex ganglion (SG6) from a normal E25 embryo. Many more labelled cells are visible here than in the ganglion shown in a. c, Sex ganglion (SG6) from an E25 embryo whose male genitalia had been ablated at E10. Comparison with the ganglia in the preceding two panels reveals the absence of the additional cell division characteristic of sex ganglia in normal animals. d, High-magnification micrograph from a normal sex SG of an E24 animal allowed to survive for 2 h after the injection of BrdUTP. Isolated pairs of labelled cells can be seen within the ganglion (arrowheads), apparently the siblings from recently divided precursor cells that had incorporated BrdU. The short survival time argues in favour of the additional dividing cells being born within the sex ganglia. e, f, High-magnification micrographs of small regions of normal sex ganglia from animals that were 6 months old. PIC neurons are identified by their small cell-body size and their labelling with FMRFamide antiserum (orange). Three PIC neurons with BrdU-labelled nuclei (yellow) can be seen in e, and one in f, which also shows several PIC neurons with unlabelled nuclei. Larger neurons also labelled with FMRFamide antiserum can also be seen in these panels. The expression of BrdU-immunoreactivity by these PIC neurons shows that they were born soon after the time of BrdU injection (E24). Bar in c, corresponds to 100  $\mu$ m for a, b and c, 15  $\mu$ m for d, and to 7.5  $\mu$ m for e and f.

**METHODS:** Administration of BrdU: Animals were anesthetized in 15 mM chlorobutanol in artificial pond water. For injection, micropipettes were filled with a solution of 30 mM BrdUTP and 0.1% Fast Green in leech Ringer's solution<sup>20</sup>. This solution (100–300  $\mu$ l) was pressure-injected into the blood sinus that surrounds the ventral nerve cord, and animals returned to artificial pond water until further processing. For BrdU incorporation *in vitro*, anaesthetized animals were opened along the dorsal midline and pinned in Sylgard-coated dishes, where their ventral nerve cords were exposed under sterile conditions. Slits were made in the surrounding blood sinus above the ganglia of interest using a sharp pin. Animals were then placed in L15 culture medium (supplemented with glucose, glutamine, fetal calf serum and gentamycin<sup>21</sup>), with BrdUTP added to a final concentration of 30  $\mu$ M. Animals were left in culture for up to 1 day at room temperature. Immunocytochemical detection of BrdU: Ganglia were exposed by opening the blood sinus that encloses the nerve cord, then treated with a collagenase solution (2,000 units  $\text{ml}^{-1}$  in leech Ringer's; Sigma type IV) to facilitate antibody penetration. Fixation was in 70% ethanol for 1–1.5 h, followed by rinsing in PBS-TX (0.01 M PBS with 1% Triton X-100). Tissue was then treated with 2 N HCl for 40 min to denature the DNA. After brief incubation in 0.1 M  $\text{Na}_2\text{B}_4\text{O}_7$  (to neutralize the acid) and a PBS wash, the tissue was incubated at 4 °C overnight with mouse monoclonal anti-BrdU sera (Becton-



Dickinson), diluted 1:20 with 2% normal sheep serum (Cappel). The following day the tissue was washed and incubated for 2 h with sheep anti-mouse sera (Cappel) diluted 1:20. Finally, the tissue was incubated for 2 h with mouse peroxidase-antiperoxidase (Sternberger-Meyer) diluted 1:100 with 2% normal sheep serum, followed by reacting peroxidase with diaminobenzidine. Ganglia were then dehydrated in ethanol, cleared in xylene and mounted whole. Simultaneous immunocytochemical detection of BrdU incorporation and FMRFamide immunoreactivity: ganglia were fixed in 4% paraformaldehyde (in 0.1 M phosphate buffer) for 30 min, washed and incubated at 4 °C overnight with FMRFamide rabbit antisera (diluted 1:1000) and 2% normal goat serum. Anti-FMRFamide antibodies were visualized using goat anti-rabbit biotinylated antibodies (diluted 1:50) and Texas Red-conjugated avidin D (15  $\mu\text{g ml}^{-1}$ ) in HEPES-buffered saline, pH 8.2; both reagents from Vector Laboratories<sup>22</sup>. As the paraformaldehyde fixation prevents the detection of anti-BrdU binding sites<sup>23,24</sup>, after the avidin D step the tissue was digested with 0.2% trypsin in 0.05 M Tris buffer with 0.1%  $\text{CaCl}_2$ , pH 7.6, for 6 min. After washing out the enzymes, the tissues were exposed to HCl and  $\text{Na}_2\text{B}_4\text{O}_7$  as described above. BrdU detection was done next. The anti-BrdU antibodies were revealed using fluorescein-conjugated anti-rabbit antibodies (Boehringer–Mannheim; diluted 1:25). Preparations were mounted in 100% glycerol and photographed with Kodak Tech Pan (ASA 100) or Ektachrome (400 ASA) film.

FIG. 2 *a*, Diagram showing a portion of the leech ventral nerve cord and the reproductive organs as seen in a young adult after partial dissection (adapted from Fig. 1, ref. 5). The segmental ganglia are numbered;  $\delta$ , penis and prostate;  $\varnothing$ , vagina; E, epididymides; o, ovaries; t, testisacs (present in the seventh through to the fifteenth body segments); arrowheads, vas deferens; arrows, branches of the root nerves of the sex ganglia innervating the genitalia. Here the penis, prostate and epididymides were ablated, but not the testisacs and the posterior part of the vas deferens. In the ablations of the female organs, the ovaries were usually incompletely ablated. Neither the testisacs nor the ovaries seem to receive innervation from the sex ganglia. Transplants of male genital primordia were incorporated readily into host embryos and contributed apparently normal tissues, including the penis, gonopore, prostate and epididymides (for details of surgical procedures, see ref. 5). *b*, Segmental differences in cell proliferation are evident in late embryonic development. The histograms show that the numbers of BrdU-labelled cells are significantly higher in the sex ganglia (SG5 and SG6) than in non-sex ganglia (SG4 and SG7) in E24 and E26 embryos, but are the same in E30 and P10 animals. E20 embryos already show significant segmental differences, but E15 embryos have similar numbers of labelled cells in sex and non-sex ganglia (see text). Data from animals injected with a 30 mM solution of BrdUTP and allowed to survive for 1 day. The means are based on the following numbers of specimens: E24,  $n=14$ ; E26,  $n=4$ ; E30,  $n=9$  and P10,  $n=3$ . Stars indicate no labelled cells found in the corresponding ganglia. *c*, The additional cell proliferation in the sex ganglia requires the presence of the male genitalia and their innervation by the sex ganglia. Histogram comparing the distributions of dividing cells for non-sex ganglia (SG4 and SG7) and sex ganglia (SG5 and SG6) from normal animals, from animals that had the female genitalia ablated (FGA), and from animals that had the male genitalia ablated (MGA) or SG6 per-



manently disconnected from the male genitalia (NC-SG6). Surgical procedures were done at E10. All animals were injected with BrdUTP at E24–25 and sacrificed 1 day later. Means based on the following numbers of specimens: Normal animals,  $n=27$ ; FGA,  $n=17$ ; MGA,  $n=23$  and NC-SG6,  $n=3$ . For *b* and *c*, the number of labelled cells was determined from whole-mount preparations stained with immunoperoxidase to reveal the incorporation of BrdU (see Methods in Fig. 1). Only stained cells within the boundaries of the ganglia were counted. Error bars indicate s.e.m.

had been cut, thus disconnecting these ganglia from the periphery. There was no significant difference between the number of labelled cells in the disconnected ( $16.5 \pm 2.3$  cells) and connected ( $15.3 \pm 2.7$  cells) sex SG, nor was any effect of nerve cutting *per se* seen in the non-sex ganglia. This result shows that the dividing cells do not migrate into the sex ganglia from the periphery, and that connection to the genitalia is not necessary during the actual time that differential mitosis takes place in the sex ganglia.

Our observations are the first to demonstrate that a peripheral target organ exerts a localized and temporally restricted control over the birth of a particular population of central neurons. A few previous reports in vertebrate<sup>2,11–14</sup> and invertebrate<sup>15–17</sup> systems have presented suggestive evidence along these lines, but the data were not conclusive. More recent experiments, in which developing gut<sup>18</sup> or skeletal and cardiac muscle<sup>19</sup> were juxtaposed with spinal cord neuroepithelium have shown a localized mitogenic effect on the cord, although innervation of the transplant is apparently not required, in contrast to what we have observed in the leech.

The elements of the connections between the sex ganglia and the male genitalia responsible for triggering the birth of the PIC neurons remain to be discovered. It is interesting that the critical

time for the interaction with the genitalia precedes the period of cell division by several days: connection to the periphery is necessary through E15, but the PIC neurons are born after this time. A possible explanation for this phenomenon is that the interaction provides a mitogenic signal up to about E15 that triggers precursors of the PIC neurons, present only in the sex SG, to begin a protracted set of divisions that are independent of the continued presence of this signal. It is also possible that the interaction provides a trophic signal that ensures the survival of these precursors, which later divide independently of this signal. The validity of these suggestions requires further experimental testing. □

Received 12 April; accepted 12 June 1990.

1. Oppenheim, R. W. *Trends Neurosci.* **12**, 252–255 (1989).
2. Williams, R. W. & Herrup, K. A. *Rev. Neurosci.* **11**, 423–453 (1988).
3. Macagno, E. R. *J. comp. Neurol.* **190**, 283–302 (1980).
4. Stewart, R. R., Spergel, D. & Macagno, E. R. *J. comp. Neurol.* **253**, 253–259 (1986).
5. Baptista, C. A. & Macagno, E. R. *J. Neurobiol.* **19**, 707–726 (1988).
6. Kuhlman, J. R., Li, C. & Calabrese, R. L. *J. Neurosci.* **5**, 2301–2309 (1985).
7. Baptista, C. A. & Macagno, E. R. *Neuron* **1**, 949–962 (1988).
8. Gratzner, H. G. *Science* **218**, 474–475 (1982).
9. Truman, J. W. & Bate, M. *Dev. Biol.* **125**, 145–157 (1988).
10. Miller, M. W. & Nowakowski, R. S. *Brain Res.* **457**, 44–52 (1988).
11. Bibb, H. D. *J. exp. Zool.* **200**, 265–276 (1977).



12. Kollros, J. J. *J. comp. Neurol.* **205**, 171-178 (1982).
13. Davis, M. R. & Constantine-Paton, M. *J. comp. Neurol.* **221**, 444-452 (1983).
14. Devor, M., Govrin-Lippman, R., Frank, I. & Raber, P. *Somatosensory Res.* **3**, 139-167 (1985).
15. Chiarodo, A. T. *J. exp. Zool.* **153**, 263-277 (1963).
16. Macagno, E. R. *Dev. Biol.* **73**, 206-238 (1979).
17. Anderson, H., Edwards, J. S. & Palka, J. A. *Rev. Neurosci.* **3**, 97-139 (1980).
18. Rothman, T. P., Gershon, M. D., Fontaine-Perus, J. C., Chanconie, M. & Le Dourain, N. M. *Dev. Biol.* **124**, 331-346 (1987).
19. Fontaine-Perus, J. C., Chanconie, M., Le Dourain, N. M., Gershon, M. D. & Rothman, T. P. *Development* **107**, 413-422 (1989).
20. Muller, K. J., Nicholls, J. G. & Stent, G. S. (eds) *Neurobiology of the Leech* (Cold Spring Harbor Laboratory, 1981).
21. Fuchs, P. A., Nicholls, J. G. & Ready, D. F. *J. Physiol.* **316**, 203-223 (1981).
22. Penit, C. *J. Histochem. Cytochem.* **36**, 473-478 (1988).
23. Harms, G., Van Goor, H., Koudstaal, J., De Ley, L. & Hardonk, M. J. *Histochemistry* **86**, 393-395 (1987).
24. Hayashi, Y., Koike, M., Matsutani, M. & Hoshino, T. *J. Histochem. Cytochem.* **36**, 511-514 (1988).

ACKNOWLEDGEMENTS. We thank K. Muller, L. Wolszon, D. Kelley and B. Passani for their critical commentaries on the manuscript, and N. Necles for help with the photography. This work was supported by the NIH and a Javits Neuroscience Award.

## Two rat homologues of *Drosophila achaete-scute* specifically expressed in neuronal precursors

Jane E. Johnson, Susan J. Birren  
& David J. Anderson\*†

Division of Biology 216-76 and \*Howard Hughes Medical Institute,  
California Institute of Technology, Pasadena, California 91125, USA

IN vertebrates, the peripheral nervous system is embryologically derived from the neural crest<sup>1</sup>. Although the earliest neural crest cells seem to be multipotent<sup>2,3</sup>, the molecular mechanisms responsible for the restriction of these cells to different sublineages<sup>4</sup> are not understood. We therefore searched for developmental control genes expressed in crest cells or their derivatives. One important class of regulatory molecules comprises proteins with common DNA-binding and dimerization domains, the basic helix-loop-helix (B-HLH) region<sup>5</sup>. Members of this family include MyoD (ref. 6), a mammalian myogenic determination molecule, and proteins encoded by genes of the achaete-scute complex<sup>7-9</sup> of *Drosophila*, which have an important role in neuronal determination<sup>10</sup>. From a sympathetic neuronal precursor cell line derived from the neural crest<sup>11</sup> we have now isolated two different mammalian genes that are homologous to genes of the achaete-scute complex. The sequence of the B-HLH-encoding region of these genes is more similar to that of the genes of the achaete-scute complex than it is to that of other, mammalian members of the B-HLH family. At least one of these genes is transiently expressed in the embryonic rat nervous system, is not detected in non-neuronal tissues or cell lines, and is induced by nerve growth factor in PC12 cells.

To isolate neural-specific genes encoding B-HLH proteins and expressed in neural crest derivatives, we amplified complementary DNA from a sympathoadrenal progenitor ('MAH') cell line<sup>11</sup>, using degenerate oligonucleotide primers and the polymerase chain reaction (PCR). Initial attempts using primers based on the sequences of mammalian B-HLH proteins were unsuccessful. Subsequently, we used primers based solely on the amino-acid sequence of the proteins encoded by genes of the achaete-scute complex (AS-C) (Fig. 1b). This experiment yielded several independent clones defining two closely related but distinct sequences (Fig. 1a). We have named these clones

*MASH-1* and *MASH-2*, for mammalian achaete-scute homologues. Screening of over 100 additional clones from this amplification produced additional isolates of *MASH-1* and *MASH-2*, but no new homologous genes.

The *MASH-1* and *MASH-2* cDNAs encode proteins of predicted relative molecular masses of roughly 25,000 and 27,700, respectively. Both proteins contain HLH and basic domains, which are 92% identical to one another (Fig. 1a, bold type) and are in turn about 80% identical to the comparable region of the *Drosophila* AS-C genes (Fig. 1b, upper), although the length of the mammalian loop is considerably smaller. Strikingly, *MASH-1* and *MASH-2* are more closely related to the AS-C genes in this region than they are to other, mammalian, members of the HLH family (Fig. 1b, lower). Outside the B-HLH region, however, the sequences of both *MASH-1* and *MASH-2* diverge completely, not only from each other, but also from those of their *Drosophila* counterparts.

To determine the size and expression pattern of messenger RNA transcripts encoded by the *MASH* genes, we examined a series of clonal cell lines representing neural crest and non-crest derivatives. *MASH-1* cDNA probes detected a transcript of ~2.8-3.0 kilobases (kb) in MAH cells<sup>11</sup> (Fig. 2, lanes 3 and 4). Expression was not detected in an MAH cell clonal variant that lacks the lineage markers characteristic of the parental cell line (Fig. 2, lane 5; S.J.B. and D.J.A., unpublished results). Expression of *MASH-1* was also detected in PC12 cells<sup>12</sup> (which derive from the same sympathoadrenal lineage as MAH cells), where it was upregulated by nerve growth factor (Fig. 4a, lanes 9 and 10). Expression was not detected in schwannoma or melanoma cell lines representing other, non-neuronal neural crest derivatives (Fig. 2, lanes 7 and 9), but was observed in a crest-derived thyroid medullary carcinoma endocrine cell line (not shown). No *MASH-1* expression was detected in rat fibroblast or liver cell lines, or in RINr cells, a pancreatic endocrine cell line of endodermal origin (Fig. 2, lanes 1, 2 and 6, respectively). Although this survey is not exhaustive, it suggests that *MASH-1* expression is restricted to cell lines of neuronal or neuroendocrine origin. The human neuroblastoma line SY5Y, however, did not express detectable *MASH-1* (Fig. 2, lane 10), suggesting either that expression is restricted to certain classes of neuronal progenitor cells, or that our rat *MASH-1* probe does not cross-hybridize to its human counterpart.

A reprobing of the northern blot of Fig. 2 (containing total RNA) with *MASH-2* probes failed to reveal transcripts in any of the cell lines examined. However, blots containing poly(A)<sup>+</sup> mRNA and hybridized with the *MASH-2* probe (Fig. 3) revealed transcripts of ~1.8 and 1.3 kb in MAH and PC12 cells, but not in Rat-1 fibroblasts (Fig. 3b, lanes 2, 3 and 4). The abundance of these transcripts is much lower than that of *MASH-1* mRNA (compare with Fig. 3a, lanes 2 and 3). This more sensitive analysis also revealed a second lower-abundance *MASH-1* transcript of ~2 kb. These data therefore suggest that *MASH-2* may also be specifically transcribed in neuronal precursors, albeit at much lower levels than *MASH-1*.

Genetic evidence indicates that the *Drosophila* AS-C genes function in the early development of both the central and peripheral nervous systems<sup>10</sup>. We therefore analysed *MASH-1* expression in embryonic rat neuronal tissues. Northern blots revealed that *MASH-1* is transiently expressed in the developing brain (Fig. 4a). Expression is first detectable at embryonic day (E)10.5-11.5, rises slightly and then decays by E20.5 (Fig. 4a, lanes 1-5). Little, if any, expression is detectable in adult brain total RNA (Fig. 4a, lane 6). The low numbers of neural crest cells in early embryos preclude northern blot analysis of this population; therefore, we do not know the earliest time at which *MASH-1* is expressed during ontogeny of the peripheral nervous system. But the levels of *MASH-1* mRNA in postnatal adrenal medulla and superior cervical sympathetic ganglion (Fig. 4a, lanes 7 and 8) were much lower than in PC12 cells (compare Fig. 4a, lane 9) or in MAH cells (compare Fig. 3a, lanes 2 and

† To whom correspondence should be addressed.

3), which represent a common precursor to neurons of the superior cervical sympathetic ganglion and chromaffin cells of the adrenal medulla<sup>11</sup>. This suggests that expression of *MASH-1* may also be transient during development of the peripheral nervous system.

The expression of *MASH-1* in midgestational embryos seems to be neural-specific. At E16.5, no signal was detectable in liver, heart or kidney (Fig. 4b, lanes 1, 3, 4). A weak signal was detected in lung (Fig. 4b, lane 2)—which contains parasympathetic ganglia at this stage of development—and also in E16.5 adrenal gland (Fig. 4b, lane 5), which contains a subpopulation (5–10%) of sympathoadrenal progenitor cells<sup>13</sup>. No expression was detected in either adult mouse skeletal muscle or heart (Fig. 4, lanes 9 and 10). No transcripts hybridizing to *MASH-2* cDNA or cRNA probes could be unambiguously identified

on the northern blots of Fig. 4, probably reflecting the lower abundance of this transcript (data not shown).

Here we have described the isolation and characterization of two mammalian homologues of genes of AS-C<sup>10</sup>. As the AS-C genes, like other members of the B-HLH family, encode DNA-binding proteins<sup>14</sup>, it seems likely that the *MASH* genes have a similar activity. Indeed, only the putative DNA-binding and dimerization domains have been conserved between the products of the *MASH* genes and their *Drosophila* counterparts. In the case of MyoD, deletion analysis has shown that this region comprises the minimal functional unit of the protein<sup>15</sup>.

*MASH* genes, like their *Drosophila* homologues, are specifically expressed in neuronal precursors. Although many mammalian homologues of *Drosophila* regulatory genes have been isolated<sup>16</sup>, in this case the sequence conservation is reflected



FIG. 1 Deduced amino-acid sequences of *MASH-1* and *MASH-2* and their relationships to other B-HLH proteins (single-letter amino-acid code). **a**, Aligned sequences of *MASH-1* and *MASH-2*. The B-HLH region is in bold type, and sequence identities surrounding this region are underlined. The predicted initiator Met in each case is the most upstream in the open reading frame set by the B-HLH region, and in the case of *MASH-1* was confirmed by *in vitro* transcription-translation experiments. No other long open reading frames were observed in the cDNAs sequenced (see below). Note the unusually long runs of alanines and glutamines N-terminal to the basic region of *MASH-1*, which are found in several other regulatory genes. **b**, Boxed regions indicate sequence identities between *MASH-1* and *MASH-2* and the four *Drosophila* AS-C genes. White letters indicate consensus residues present in most or all members of the HLH family<sup>21</sup>. Note that the sequences in between these consensus residues are far more conserved between the products of the *MASH* and AS-C genes than between these products and other members of the HLH family. A selected subset of these other HLH sequences are shown, each representing different subfamilies that have been identified. The residues illustrated from the products of the following genes are: AS-C T5, 16–98; T8, 86–165; T4, 91–169; T3, 75–153 (refs 7, 8). The sequences of the gene products of Ly1-1, c-Myc, MyoD1, da,

E47 and twist are taken from Fig. 5 in ref. 21; e(sp) corresponds to the product of the *enhancer of split* gene *m8*, residues 9–65 (ref. 22); *hairy* gene product residues are 30–88 (ref. 23).

**METHODS.** Random-primed cDNA made from MAH cell<sup>11</sup> total cytoplasmic RNA was subjected to 35 cycles of amplification by PCR, at an annealing temperature of 45 °C, using fully-degenerate primers corresponding to the regions underlined by arrows in **b**. Fourfold degenerate residues were substituted by inosine and restriction sites *EcoRI* or *BamHI* were included at the 5' end of each primer. The bands of interest were excised from a 4% Nu-Sieve agarose gel, and subcloned into pGEM3. Clones containing *MASH* cDNAs were identified by sequencing. Of 24 clones initially sequenced, 16 corresponded to *MASH-1* and two to *MASH-2*. Probes derived from these amplified cDNAs were used to screen MAH and PC12 cDNA libraries. Longer clones isolated from these libraries were used to confirm and extend the sequence obtained from the original PCR-amplified cDNAs. The precise relationship of these longer cDNAs to the transcripts detected on northern blots has not yet been determined. *MASH-1* and *MASH-2* cDNAs (1,354 and 1,440 nucleotides, respectively) were sequenced in their entirety on both strands. These nucleotide sequences are available from the EMBL database, accession nos X53724 (*MASH-2*) and X53725 (*MASH-1*).



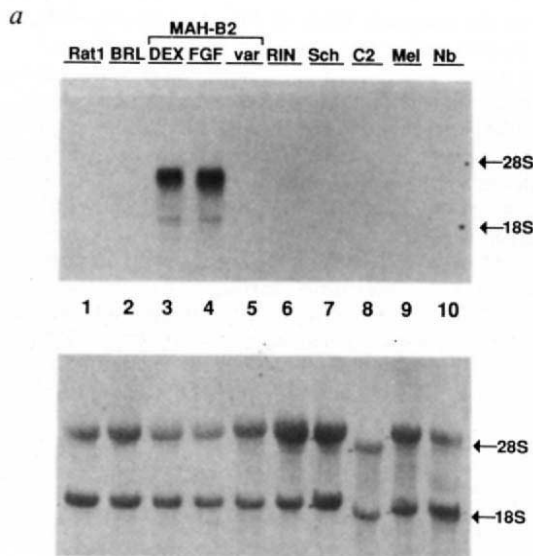


FIG. 2  
fibroblast  
genitor  
grown  
growth  
that he  
D.J.A.,  
rat schwannoma line; C2, mouse C2C12 muscle cells; Mel, human A875 melanoma cells; Nb, human SY5Y neuroblastoma cells. In another experiment, a relatively high level of expression was detected in RNA from a rat thyroid medullary carcinoma (TMC) cell line (data not shown). The faint band running just ahead of the 18S ribosomal RNA marker in lanes 3 and 4 represents a second, lower-abundance *MASH-1* transcript that is more clearly seen on northern blots using poly(A)<sup>+</sup> RNA (Fig. 3).  
METHODS. Total RNA (15 µg) isolated from the cell lines indicated was run in each lane. The RNA was fractionated on a 1% formaldehyde-agarose gel and transferred to a nylon membrane (GeneScreen). The northern blot was probed with a randomly-primed 0.3-kb *PvuII* fragment located 3' to the HLH region in a *MASH-1* cDNA. The blot was washed to high stringency (0.5 × SSC buffer, 1% SDS, 68 °C), and exposed for three days. Equal amounts of intact RNA were run in each lane, as indicated by the intensities of the 18S and 28S ribosomal RNA bands revealed by methylene blue staining of the filter before hybridization (b).

FIG. 4 Transient and specific expression of *MASH-1* mRNA in the developing rat central nervous system. a, E11–20 indicate approximate gestational ages of fetal rat brain. Ad, adult brain tissue. SCG, superior cervical sympathetic ganglion (neonatal). AM, adrenal medulla (adult). PC12<sup>−</sup> and PC12<sup>+</sup> indicate PC12 cells treated without or with 50 ng ml<sup>−1</sup> β nerve growth factor for two days, respectively. The faint smudge below the band of interest in lanes 1–4 represents residual hybridization to 18S rRNA from an unrelated probe used in a previous hybridization. b, Northern blot of total RNA from E16.5 tissues: liver (Liv), lung (Lng), heart (Hrt), kidney (Kid), adrenal gland (Adr) and brain (Br). For a direct comparison of signal intensities, RNA samples from E12 brain and NGF-treated PC12 cells (PC<sup>+</sup>) were run in parallel; these correspond to the samples shown in a, lanes 2 and 10, respectively. Note the low level of *MASH-1* expression in adrenal gland. Lanes 9 and 10 contain RNA from adult mouse skeletal muscle (skm) and heart (hrt).  
METHODS. Total cellular RNA (15 µg) was run in each lane. Conditions of transfer and hybridization were as in Fig. 2. Equal amounts of intact RNA were run in each lane, as indicated by the intensities of rRNA bands revealed by either ethidium bromide staining of the gel before transfer (a, lower panel), or methylene blue staining of the filter before hybridization (b, lower panel).

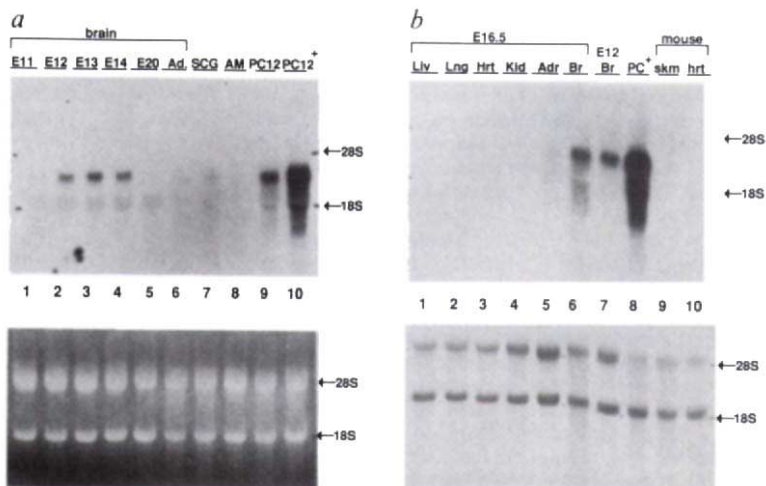


FIG. 3 Detection of *MASH-2* mRNAs in MAH and PC12 cells. Duplicate northern blots containing poly(A)<sup>+</sup> RNA from MAH, PC12 and Rat-1 fibroblast cell lines were probed with either *MASH-1* (a) or *MASH-2* (b) cDNA probes. The probes were derived from regions located 3' to the HLH-basic region that do not cross-hybridize. Similar amounts of probe were used in both cases, and the filters exposed to film for 4.5 days with an intensifying screen. Note the absence of expression of both *MASH-1* and *MASH-2* in Rat-1 fibroblasts. Note also that *MASH-2* transcripts are virtually undetectable in MAH cell total RNA (b, lane 1).  
METHODS. Either 5 µg (lanes 1) of total RNA or 3 µg of poly(A)<sup>+</sup> RNA, isolated using the Fast Track Kit (In Vitrogen, Inc.), (lanes 2–4) from the indicated cell lines were run in each lane. Conditions of transfer and hybridization were as in Fig. 2. Equal amounts of poly(A)<sup>+</sup> RNA were run in each lane, as indicated by the intensities of residual 18S rRNA bands revealed by methylene blue staining of the filter before hybridization (lower panels).

in a conservation of tissue-specificity. A similar observation has recently been made for a *Drosophila* homologue of *MyoD*, which is specifically expressed in *Drosophila* myogenic precursors (A. Michelson, S. Abmayr and T. Maniatis, personal communication). Like the *MASH* and AS-C genes, the *Drosophila* and mammalian *MyoD* genes are more closely related to each other than they are to other B-HLH-encoding genes in the same species. This parallel conservation of lineage-specificity and sequence may indicate invariant amino-acid residues required for the determination of a particular cell type; such residues have been identified in the basic region of MyoD<sup>17</sup>. Mammalian and *Drosophila* AS-C genes likewise share a pattern of conserved residues in the basic region as well as a three-residue gap (VARR...N(E/A)RERNRVK; single-letter amino-acid code), not found in other members of the B-HLH gene family (Fig. 1b).

The expression pattern of *MASH-1* has some broad similarities to that of the AS-C genes in *Drosophila*. For example, the AS-C genes control the development of both central and peripheral neurons<sup>18</sup>, and our data indicate that *MASH-1* is expressed in both a peripheral sympathoadrenal progenitor as well as in some central nervous system precursors. In addition *MASH-1*, like its *Drosophila* counterparts<sup>19</sup>, is transiently expressed during neuronal development. These data, together with the sequence conservation, suggest that the *MASH* genes may function in a manner broadly similar to their fly homologues. This idea is supported by the finding of several *MASH* genes, as *Drosophila* contains at least four related AS-C genes. In the fly, these genes control the development of different classes of peripheral sensory neurons<sup>10</sup>, suggesting that the two *MASH* genes could similarly function in different neuronal subtypes. Unfortunately, transfection of *MASH-1* expression constructs into different cell lines (for example, NIH3T3, PC12, HeLa, 10T1/2) has so far failed to reveal any detectable phenotypic changes, as assayed by morphology, expression of neuronal-specific markers, or transactivation of chloramphenicol acetyltransferase gene constructs containing neural-specific regulatory regions<sup>20</sup>. The function of *MASH* genes in neuronal development therefore remains to be established. Nevertheless, our observations indicate that some aspects of neurogenesis in flies and mammals may be more similar than previously anticipated. □

## Distinct sequence of negative or positive selection implied by thymocyte T-cell receptor densities

Pamela S. Ohashi, Hanspeter Pircher, Kurt Bürki\*, Rolf M. Zinkernagel & Hans Hengartner

Institute of Pathology, Department of Experimental Pathology, University Hospital, 8091 Zürich, Switzerland

\* Preclinical Research, Sandoz Pharma Limited, 4002 Basel, Switzerland

RECENT evidence suggests that positive and negative selection of thymocytes bearing  $\alpha\beta$  T-cell receptors occurs during the predominant double-positive ( $CD4^+CD8^+$ ) stage<sup>1-13</sup>. But the sequence or stage at which positive or negative selection occurs during thymocyte maturation has not been well defined<sup>14,15</sup>. Here we use transgenic mice to show that the  $CD4^+CD8^+$  stage might be further subdivided into  $CD3^{lo}$  (low) and  $CD3^{in}$  (intermediate) stages. The  $CD3^{in}$  stage could represent T cells that have been positively selected, as this stage is dependent on the presence of the appropriate major histocompatibility complex restriction element. In addition, we use two different tolerizing antigens to show that negative selection may occur either before or after this  $CD3^{in}$  stage.

Transgenic mice carrying the T-cell receptor (TCR)  $\alpha\beta$  chains (V $\alpha$ 2-J $\alpha$ TA31, V $\beta$ 8.1DJ $\beta$ 2.4) specific for the lymphocytic choriomeningitis virus (LCMV) presented in the context of H-2<sup>b</sup> have been generated<sup>13</sup>. Positive selection of the T cells bearing the transgenic  $\alpha\beta$  TCR is dependent on the presence of the H-2D<sup>b</sup> molecule. This results in the skewing of both thymocytes and peripheral T cells to the CD8 lineage<sup>13,16</sup>.

Thymocytes from the  $\alpha\beta$  TCR transgenic line 327 have been compared in H-2<sup>bb</sup>, H-2<sup>bd</sup> and H-2<sup>dd</sup> haplotypes to examine the effects of the major histocompatibility complex (MHC) on selection of transgenic thymocytes. In C57BL/6 control mice, thymocytes stained with a monoclonal antibody against CD3 (ref. 17) show the characteristic biphasic distribution, including the major  $CD3^{in}$ - $CD3^{lo}$  population and a  $CD3^{hi}$  population (Fig. 1a)<sup>18,19</sup>. Cytofluorometric analysis of thymocytes from H-2<sup>bb</sup> or H-2<sup>bd</sup> transgenic mice reveal three populations on the basis of their level of CD3 expression; a  $CD3^{lo}$  population (33%), an unusual  $CD3^{in}$  phenotype (43%) and a  $CD3^{hi}$  population (24%) (Fig. 1c, d). The same profile is seen in transgenic H-2<sup>bb</sup> and H-2<sup>bd</sup> mice. This indicates that the amount of H-2<sup>b</sup> expressed on the surface does not alter the CD3 profile and that H-2<sup>d</sup> does not have a negative influence on the transgenic TCR population. Thymocytes from an H-2<sup>dd</sup> transgenic mouse have a  $CD3^{lo}$  peak that is a little more intense than the  $CD3^{lo}$  peak in the control C57BL/6 mouse (Fig. 1a, b). The slight shift in the  $CD3^{lo}$  peak is probably due to the immediate expression of the functional transgenic TCR in most thymocytes. The H-2<sup>dd</sup> transgenic mouse has a  $CD3^{hi}$  density peak that is similar to the normal thymus profile. But the H-2<sup>dd</sup> transgenic mouse does not have the  $CD3^{in}$  peak, suggesting that the presence of H-2<sup>b</sup> results in the  $CD3^{in}$  peak.

Staining the transgenic thymocytes with the V $\beta$ 8-specific monoclonal antibody KJ16 (ref. 20) gives results similar to the CD3 profiles in all transgenic mice (Fig. 1e-h). This confirms that these CD3 profiles reflect the transgenic TCR population. The H-2<sup>dd</sup> transgenic animals consistently show a relatively high frequency of KJ16 reactive cells. Using a monoclonal antibody specific for the transgenic  $\alpha$  chain (V $\alpha$ 2), data indicate that V $\alpha$ 2 is not expressed at high levels in the transgenic H-2<sup>dd</sup> animals (H.P. *et al.*, unpublished results), suggesting that the endogenous  $\alpha$  chain together with the transgenic  $\beta$  chain can pair and undergo positive selection. Taken together, these data suggest that the transgenic TCR-CD3 complex is expressed on the

Received 10 May; accepted 21 June 1990.

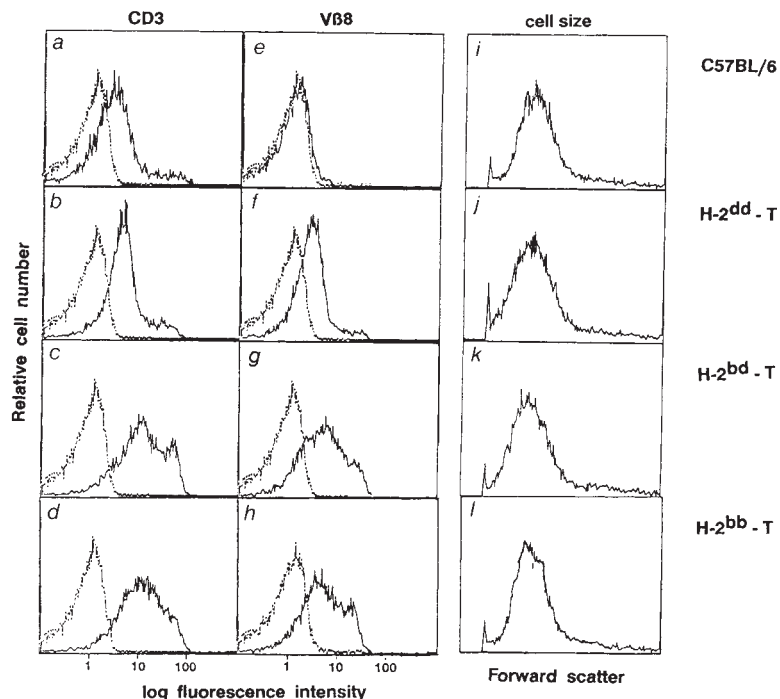
1. Le Douarin, N. M. *The Neural Crest* (Cambridge University Press, Cambridge, UK, 1982).
2. Sieber-Blum, M. & Cohen, A. M. *Devl Biol.* **80**, 96-106 (1980).
3. Bronner-Fraser, M. & Fraser, S. E. *Nature* **335**, 161-164 (1988).
4. Anderson, D. J. *Neuron* **3**, 1-12 (1989).
5. Murre, C., Schonleber-McCaw, P. & Baltimore, D. *Cell* **56**, 777-783 (1989).
6. Davis, R. L., Weintraub, H. & Lassar, A. B. *Cell* **51**, 987-1000 (1987).
7. Villares, R. & Cabrera, C. V. *Cell* **50**, 415-424 (1987).
8. Alonso, M. C. & Cabrera, C. V. *EMBO J.* **7**, 2585-2591 (1988).
9. Gonzalez, F., Romani, S., Cubas, P., Modolelli, J. & Campuzano, S. *EMBO J.* **8**, 3553-3562 (1989).
10. Dambly-Chaudière, C. & Ghysen, A. *Genes Dev.* **1**, 297-306 (1987).
11. Birren, S. J. & Anderson, D. J. *Neuron* **4**, 189-201 (1990).
12. Greene, L. A. & Tischler, A. S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2424-2428 (1976).
13. Anderson, D. J. & Axel, R. *Cell* **47**, 1079-1090 (1986).
14. Murre, C. *et al. Cell* **58**, 537-544 (1989).
15. Tapscott, S. J. *et al. Science* **242**, 405-411 (1988).
16. Graham, A., Papalopulu, N. & Krangel, R. *Cell* **57**, 367-378 (1989).
17. Davis, R. L., Cheng, P. F., Lassar, A. B. & Weintraub, H. *Cell* **60**, 733-746 (1990).
18. Ghysen, A. & Dambly-Chaudière, C. *Trends Genet.* **5**, 251-255 (1989).
19. Cabrera, C. V., Martínez-Arias, A. & Bate, M. *Cell* **50**, 425-433 (1987).
20. Mori, N., Stein, R., Sigmund, O. & Anderson, D. J. *Neuron* **4**, 583-594 (1990).
21. Mellentin, J. D., Smith, S. D. & Cleary, M. L. *Cell* **58**, 77-83 (1989).
22. Klambt, C., Knust, E., Tietze, K. & Campos-Ortega, J. A. *EMBO J.* **8**, 203-210 (1989).
23. Rushlow, C. A. *et al. EMBO J.* **8**, 3095-3103 (1989).

ACKNOWLEDGEMENTS. We thank K. Zinn, D. Stemple and C. Schoenherr for independently suggesting the use of AS-C-biased probes; T. Wilkie for help with PCR cloning; M. Palazzolo and M. Strathman for advice on cDNA library construction; C. Puckett for a PC12 cDNA library; B. Wold and colleagues, Dr R. Axel and P. Patterson for discussion; Dr T. Maniatis for communicating his results before publication; and S. M. Padilla for technical assistance. J.E.J. and S.J.B. are supported by postdoctoral fellowships from the Muscular Dystrophy Association and Damon Runyan-Walter Winchell Cancer Fund, respectively. This work was supported by a Searle Scholars Award, an NSF PY1 Award and a Sloan Foundation Faculty Research Fellowship (D.J.A.). D.J.A. is an Assistant Investigator of the Howard Hughes Medical Institute.



FIG. 1 TCR-CD3 density varies with the MHC antigens present during ontogeny. Thymocytes from C57BL/6 control mice, H-2<sup>bb</sup>, H-2<sup>bd</sup>, or H-2<sup>dd</sup> transgenic mice were stained with monoclonal antibodies to CD3 (ref. 17), (a-d) or Vβ8 (ref. 20) (e-h). Broken lines indicate control thymocytes stained with fluorescein-conjugated goat anti-rat antibodies. Analysis of thymocyte size using the forward light scatter is also shown (i-l).

METHODS. Single cell suspensions from the thymuses of various mice were prepared and  $1 \times 10^6$  thymocytes were stained with the given antibody in BSS, 1% BSA and 0.1% sodium azide. Cells were incubated for 20 min on ice, then washed. Viable cells (10,000) were analysed from each sample. Transgenic animals were backcrossed to C57BL/6 mice for at least three generations and selected for H-2<sup>bb</sup> Mls<sup>b</sup> phenotype. H-2<sup>bd</sup> and H-2<sup>dd</sup> transgenic mice were obtained by breeding with BALB/c mice and typing for the presence of H-2<sup>d</sup>. At least 10 animals from each group were analysed, and representative FCM profiles were chosen.



surface of thymocytes initially at a low density, as seen in the H-2<sup>dd</sup> and the H-2<sup>bd</sup> haplotype. If the TCR is able to interact with self MHC in a positive manner, the receptor density is upregulated to an intermediate level as seen in the transgenic mice bearing H-2<sup>b</sup>. Because most T cells in transgenic H-2<sup>b</sup> mice are positively selected, an increased number of CD3<sup>hi</sup> mature cells are present in the thymus. There have not been similar analyses of other transgenic systems, and it remains to be seen whether other transgenic or nontransgenic models will support our findings.

Further studies of the transgenic thymocytes indicated that the shifts in the TCR-CD3 densities were not due to changes in cell size, as shown by the histograms comparing the forward light scatter (Fig. 1i-l). Thymocytes stained with antibodies against CD3 and CD4, CD8 indicated that the TCR-CD3<sup>in</sup> stage

does not represent an unusual double-negative population but corresponds to the double-positive phenotype seen in the control C57BL/6 mice (Fig. 2).

Because our data indicate that the TCR-CD3<sup>in</sup> phenotype could represent cells that have undergone positive selection, we investigated whether tolerance occurred before or after this stage. Previous studies using Vβ8-specific monoclonal antibodies demonstrated that T cells expressing Vβ8.1 were reactive with Mls<sup>a</sup> and were clonally eliminated during thymocyte ontogeny<sup>9,11-13</sup>. As the transgenic TCR is Vβ8.1, tolerance to Mls<sup>a</sup> was examined by analysing thymocytes from either Mls<sup>b</sup> H-2<sup>bd</sup>, Mls<sup>b</sup> H-2<sup>dd</sup>, or Mls<sup>a</sup> H-2<sup>bd</sup> transgenic mice with CD3 or KJ16. When Mls<sup>b</sup> H-2<sup>dd</sup> mice are compared with Mls<sup>a</sup> H-2<sup>bd</sup> mice, the main peak in the Mls<sup>a</sup> thymocytes express TCR-CD3 at a higher density than the major peak in the Mls<sup>b</sup> H-2<sup>dd</sup> mice (Fig. 3a, c, d; Table 1). This suggests that most of the cells in Mls<sup>a</sup> mice have been positively selected and have increased the TCR-CD3 receptor density. When the Mls<sup>a</sup> mouse is compared with the Mls<sup>b</sup> H-2<sup>bd</sup> transgenic mouse, the main TCR-CD3 peak clearly overlaps with the CD3<sup>in</sup> peak defined in the H-2<sup>bd</sup>

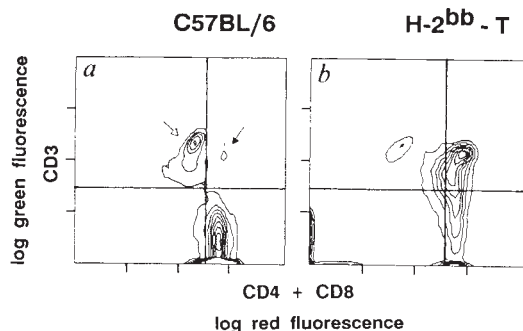


FIG. 2 TCR-CD3<sup>in</sup> stage occur on double positive thymocytes. Thymocytes from control C57BL/6 mice (a) H-2<sup>bb</sup> transgenic animals (b) were analysed by two-colour immunofluorescence. Staining with antibodies to CD3 is shown on the green axis and both CD4 and CD8 fluorescence is shown on the red axis. It should be noted that anti-CD8 antibodies stain at a higher fluorescence intensity than anti-CD4 antibodies. Open arrow single CD4-positive thymocytes; solid arrow, the CD8 single-positive population.

METHODS. Thymocytes were stained with KT3 (anti-CD3)<sup>17</sup> followed by fluorescein isothiocyanate-conjugated goat anti-rat antibodies. Cells were subsequently stained with phycoerythrin-conjugated anti-CD4 (Becton Dickinson) and biotin-conjugated anti-CD8 (Becton Dickinson), then streptavidin-conjugated phycoerythrin (Becton Dickinson) using conditions described in Fig. 1. Viable cells (50,000) were counted from these samples.

TABLE 1 Summary of thymocyte profiles

| Sample  | CD3 <sup>lo</sup> |      | CD3 <sup>in</sup> |      | CD3 <sup>hi</sup> |      |
|---------|-------------------|------|-------------------|------|-------------------|------|
|         | Peak              | Mean | Peak              | Mean | Peak              | Mean |
| Control |                   |      |                   |      |                   |      |
| C57BL/6 | 82                | 82   | —                 | —    | 155               | 156  |
| b-bd-T  | 94                | 81   | 121               | 121  | 157               | 156  |
| b-dd-T  | 94                | 96   | —                 | —    | 157               | 153  |
| a-bd-T  | ND                | ND   | 120               | 116  | ND                | ND   |
| car     | 81                | 79   | —                 | —    | 149               | 151  |
| b-car-T | 88                | 88   | —                 | —    | 167               | 164  |
| d-car-T | 96                | 92   | —                 | —    | 143               | 155  |

Peak, peak position; Mean, mean of the peak; b-bd-T, Mls<sup>b</sup> H-2<sup>bd</sup> transgenic; b-dd-T, Mls<sup>b</sup> H-2<sup>dd</sup> transgenic; a-bd-T, Mls<sup>a</sup> H-2<sup>bd</sup> transgenic; car, H-2<sup>bb</sup> nontransgenic carrier littermate; b-car-T, H-2<sup>bb</sup> transgenic carrier; d-car-T, H-2<sup>dd</sup> transgenic carrier; —, no peak present; ND, no discrete peak. The peak position and mean of the peaks were calculated using the cytologic program, from thymocytes stained with a CD3 monoclonal antibody. The peak position for thymocytes stained with goat anti-rat antibody alone is 75, and the mean is 67.

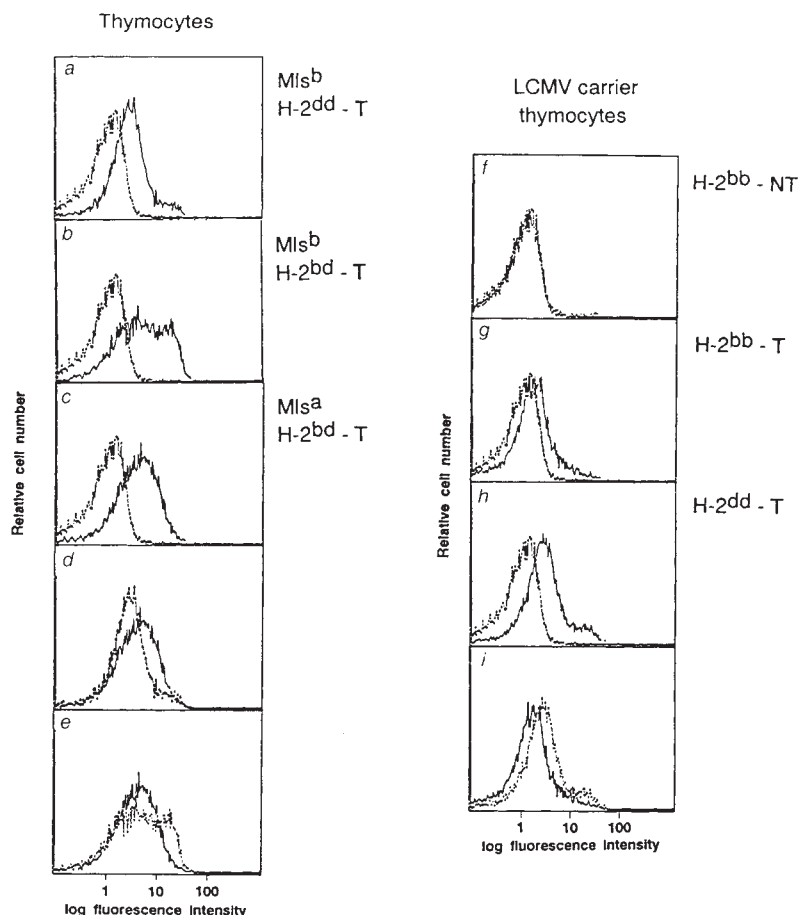


FIG. 3 Negative selection may occur at different stages of T-cell ontogeny. Thymocytes from various mice were stained with KJ16 as previously described in Fig. 1. *a-c*, Tolerance to the Mls antigen in *a*, Mls<sup>b</sup>, H-2<sup>dd</sup> transgenic (T); *b*, Mls<sup>b</sup> H-2<sup>bd</sup>-T; *c*, Mls<sup>a</sup> H-2<sup>bd</sup>-T mouse. Broken lines in *a-c* represent thymocytes stained with fluorescein-conjugated goat anti-rat antibodies. *d*, Comparison of an Mls<sup>a</sup> H-2<sup>bd</sup>-T (solid line) with an Mls<sup>b</sup> H-2<sup>dd</sup>-T mouse (broken line). *e*, Comparison of Mls<sup>a</sup> H-2<sup>bd</sup>-T (solid line) with an Mls<sup>b</sup> H-2<sup>bd</sup>-T mouse (broken line). *f-i*, Tolerance to LCMV in LCMV carrier mice that were neonatally infected with the virus. Thymocytes from an H-2<sup>bb</sup> nontransgenic (NT) carrier mouse (*f*), and an H-2<sup>bb</sup>-T (*g*) and H-2<sup>dd</sup>-T (*h*) mouse. Broken lines in *f-h* represent thymocytes stained with FITC-conjugated goat anti-rat antibodies. *i*, Composite figure showing H-2<sup>bb</sup>-T (solid line) and H-2<sup>dd</sup>-T thymocytes (broken line).

**METHODS.** Mls<sup>b</sup> H-2<sup>dd</sup> transgenic animals have been derived from matings with BALB/c animals. The Mls<sup>a</sup> H-2<sup>bd</sup> transgenic animals were F1 animals from breeding transgenic H-2<sup>bb</sup>, Mls<sup>b</sup> animals with DBA-2 mice (H-2<sup>d</sup>, Mls<sup>b</sup>). These figures are representative for at least five animals. Conditions for staining as described in Fig. 1. LCMV carrier mice were generated by neonatal LCMV infection as described previously<sup>13</sup>.

transgenic mouse (Fig. 3*b, c, e*; Table 1). The CD3<sup>hi</sup> peak in the Mls<sup>a</sup> mouse is smaller than that in the Mls<sup>b</sup> H-2<sup>bd</sup> mouse, indicating that negative selection has resulted in the reduction of the mature T-cell subset (Fig. 3*e*). The peak position and the mean of these peaks have been calculated and the data are presented in Table 1. Thus, tolerance induction to Mls<sup>a</sup> occurs after the upregulation of CD3 on transgenic thymocytes.

Tolerance to the original transgenic TCR specificity (LCMV; H-2D<sup>b</sup>) was also examined in virus carrier mice induced by neonatal infection with LCMV. The thymus from transgenic LCMV carrier mice that were either H-2<sup>bb</sup> or H-2<sup>dd</sup> were compared with nontransgenic H-2<sup>bb</sup> carrier mice. Thymuses from H-2<sup>bb</sup> transgenic carrier mice were one-tenth the size of the thymuses of carrier H-2<sup>bb</sup> littermate controls or H-2<sup>dd</sup> transgenic mice. This decrease in cell number is probably due to clonal deletion of the transgenic thymocytes as previously reported<sup>13</sup>. The H-2<sup>bb</sup> transgenic carrier mice expressed a slightly elevated CD3<sup>lo</sup> peak compared with the control nontransgenic littermate, but did not express the CD3<sup>in</sup> phenotype (Fig. 3*f, g*). When the H-2<sup>bb</sup> and H-2<sup>dd</sup> transgenic carrier mice are compared, the CD3<sup>lo</sup> or KJ16<sup>lo</sup> receptor density is lower in the H-2<sup>bb</sup> mice than in the H-2<sup>dd</sup> mice (Fig. 3*g-i*; Table 1). Taken together, these data suggest that the transgenic H-2<sup>bb</sup> carrier mouse is only allowed to express very low levels of the transgenic TCR. Therefore it seems that negative selection of LCMV occurs before the TCR-CD3<sup>in</sup> stage.

Our findings may be comparable to earlier results demonstrating the presence of cells carrying the CD4, CD8 and CD3<sup>hi</sup> antigens in human thymocyte populations. This may be related to the CD3<sup>in</sup> stage defined in our transgenic system. About 7% of the thymocytes are included in this population, and may reflect the small percentage of thymocytes that have been able to undergo positive selection<sup>21</sup>. In addition, *in vivo* administration of antibodies against CD4 results in both an upregulation of CD3 thymocyte expression, and a deletion of single positive

CD4 subset<sup>22</sup>. Signalling in this manner could mimic the signal given by 'positive selection', resulting in the increase of TCR-CD3 (refs 22, 23).

Data presented here suggest that positive selection correlates with an increase of the thymocyte TCR-CD3 density. Using this to define an intermediate stage in thymocyte ontogeny, we have shown that tolerance to a 'classical' MHC restricted antigen LCMV occurs before positive selection, whereas negative selection for Mls<sup>a</sup>-reactive T cells occurs after the TCR-CD3<sup>in</sup> stage. □

Received 2 April; accepted 25 June 1990.

- Schwartz, R. H. *Cell* **57**, 1073-1081 (1989).
- Rothenberg, E. V. *Immun. Today* **11**, 116-119 (1990).
- Marrack, P. & Kappler, J. W. *Immun. Today* **9**, 308-315 (1988).
- Teh, H. S. *et al. Nature* **335**, 229-233 (1988).
- Sha, W. C. *et al. Nature* **336**, 73-76 (1988).
- Zuniga Pflucker, J. C., Longo, D. L. & Kruisbeek, A. M. *Nature* **338**, 76-78 (1989).
- Kisielow, P., Blüthmann, H., Staerz, U. D., Steinmetz, M. & von Boehmer, H. *Nature* **333**, 742-746 (1988).
- Kappler, J. W., Roehm, N. & Marrack, P. *Cell* **49**, 273-280 (1987).
- Kappler, J. W., Staerz, U. D., White, J. & Marrack, P. *Nature* **332**, 35-40 (1988).
- MacDonald, H. R. *et al. Nature* **332**, 40-45 (1988).
- White, J. *et al. Cell* **56**, 27-35 (1989).
- Pircher, H. P. *et al. EMBO J.* **8**, 719-727 (1989).
- Pircher, H. P., Bürki, K., Lang, R., Hengartner, H. & Zinkernagel, R. *Nature* **342**, 559-561 (1989).
- Zepp, F. & Staerz, U. D. *Nature* **336**, 473-475 (1988).
- Benoist, C. & Mathis, D. *Cell* **58**, 1027-1033 (1989).
- Ohashi, P. S. *et al. Sem. Immun.* **1**, 95-104 (1989).
- Tomonari, K. *Immunogenetics* **28**, 455-458 (1988).
- Roehm, N. *et al. Cell* **38**, 577-584 (1984).
- Havran, W. L. *et al. Nature* **300**, 170-173 (1987).
- Haskins, K. *et al. J. exp. Med.* **160**, 452-471 (1984).
- Blue, M. L., Daley, J. F., Levine, H., Craig, K. A. & Schlossman, S. F. *J. Immun.* **139**, 1065-1069 (1987).
- McCarthy, S. A., Kruisbeek, A. M., Uppenkamp, I. K., Sharrow, S. O. & Singer, A. *Nature* **336**, 76-79 (1988).
- Zuniga Pflucker, J. C. *et al. J. exp. Med.* **169**, 2085-2096 (1989).

**ACKNOWLEDGEMENTS.** We thank R. MacDonald, T. Aebischer, Patrice Hugo and S. Oehen for helpful discussions and comments and M. Condrea and M. Zimmermann for flowcytometry assistance. P.S.O. is supported by the Medical Research Council of Canada. This work was funded by the Swiss NSF.



# Radial extension of macrophage tubular lysosomes supported by kinesin

Peter J. Hollenbeck & Joel A. Swanson

Department of Anatomy and Cellular Biology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA

THE centrifugal elongation of membranes to form extended tubular structures is a widespread form of intracellular organelle movement. Tubular lysosomes<sup>1</sup> and the endoplasmic reticulum<sup>2</sup>, for example, undergo such extension in association with microtubules, and this process has been mimicked *in vitro* by combining purified microtubules with isolated membranes and the mechanochemical ATPase kinesin<sup>3,4</sup>. This, along with evidence that kinesin is associated with the endoplasmic reticulum<sup>5</sup>, has led to the suggestion that kinesin provides the motive force for the formation and maintenance of elongated tubulovesicular structures in cells<sup>6,7</sup>. We have addressed this hypothesis in murine macrophages, which have prominent tubular lysosomes whose form depends on the integrity of microtubules. Here we report that two antikinesin antibodies which disrupt *in vitro* motility will each cause centripetal collapse of the array of tubular lysosomes when scrape-loaded into macrophages. To our knowledge this provides the first *in vivo* evidence that kinesin is responsible for extension of tubulovesicular structures along microtubules.

Lysosomes undergo bidirectional movement in association with microtubules<sup>8,9</sup>. In macrophages grown under appropriate conditions, the lysosomal compartment comprises an array of membrane tubules extending from the perinuclear region to the cell periphery<sup>1</sup>. Ultrastructural analysis has shown that macrophage tubular lysosomes are closely associated with microtubules, and that depolymerization of microtubules results in the centripetal collapse of the lysosomes into a perinuclear mass<sup>1</sup>. The microtubule-dependent extension of tubular lysosomes is centrifugal, and thus directed toward the plus ends of the microtubules. As the mechanochemical ATPase kinesin<sup>6,7,10-12</sup> drives movement of membrane vesicles toward the plus ends of microtubules *in vitro*<sup>13</sup>, we reasoned that it could be responsible for the generation and maintenance of the tubular lysosomal array. We tested this hypothesis by introducing into macrophages antibodies directed against kinesin and observing their effects on the distribution of tubular lysosomes.

We used two probes directed against the kinesin heavy chain in this study—polyclonal IgG purified from an antiserum against chick brain kinesin<sup>5</sup>, and SUK4, a monoclonal IgG produced against sea urchin egg kinesin and which cross-reacts with kinesin from diverse sources<sup>14</sup>. The polyclonal IgG was tested as a functional probe by applying it to kinesin-coated glass substrata and measuring its ability to inhibit subsequent ATP-dependent microtubule gliding over the surface; it caused a concentration-dependent inhibition of the percentage of microtubules exhibiting gliding motility (Fig. 1a). On immunoblots of macrophage protein, the polyclonal IgG recognized a single polypeptide with an estimated relative molecular mass of 117,000 ( $M_r$  117K) (Fig. 1b). The SUK4 monoclonal anti-kinesin, which has previously been shown to inhibit kinesin-based motility *in vitro*<sup>14</sup> by binding to an epitope on the globular head of the molecule<sup>15</sup>, also recognized a single 117K band on immunoblots; we presume this to be the macrophage kinesin heavy chain (Fig. 1b). Both probes thus seemed likely to disrupt kinesin-based functions in macrophages.

To measure the effects of anti-kinesin on the extension of tubular lysosomes, we first fluorescently labelled the lysosomal compartment by allowing macrophages to pinocytose ovalbumin conjugated to the fluorochrome Texas Red. We then introduced polyclonal or monoclonal anti-kinesin IgGs into the

macrophages by the method of scrape-loading<sup>16</sup>. By including in the scrape-loading buffer a fluorescein-conjugated dextran with a Stokes' radius similar to that of immunoglobulin, we could identify by their fluorescence cells that had been successfully loaded with immunoglobulin. As an alternative, in some experiments immunoglobulin was loaded into cells without an accompanying fluorescent dextran marker, cells were then fixed and those that had been successfully loaded were detected by positive immunostaining with fluorescently labelled antibodies against the scrape-loaded IgGs. This procedure allowed larger numbers of cells to be examined, and also revealed the location of anti-kinesin in the scrape-loaded cells. As a control for the possible nonspecific effects of antibody loading, we examined cells loaded with IgG purified from preimmune serum. We also examined cells loaded with fluorescein-conjugated dextran alone to control for the effects of the scrape-loading procedure itself.

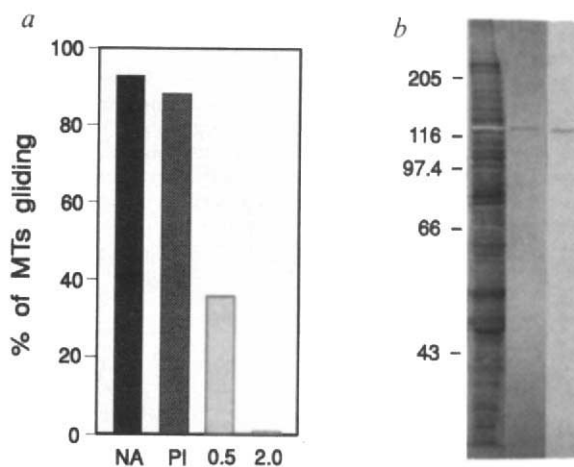


FIG. 1 Characterization of anti-kinesin antibodies as functional probes. *a*, ATP-dependent gliding of microtubules (MTs) over kinesin-coated substrata was inhibited by polyclonal anti-kinesin IgG, but not by IgG from preimmune serum. NA, no antibody treatment; PI, preimmune IgG, 2.0 mg ml<sup>-1</sup>; 0.5, 0.5 mg ml<sup>-1</sup> anti-kinesin IgG; 2.0, 2.0 mg ml<sup>-1</sup> anti-kinesin IgG. For each treatment, 25–50 microtubules were scored in several fields of view. *b*, Recognition of the macrophage kinesin heavy chain by both polyclonal anti-kinesin and SUK4 on immunoblots. Macrophage protein (8 µg) was electrophoresed on a 9% SDS-polyacrylamide gel, transferred to nitrocellulose, and immunostained with polyclonal or monoclonal anti-kinesin. Protein staining (left), polyclonal anti-kinesin immunostaining (centre) and SUK4 monoclonal anti-kinesin immunostaining (right) of identical lanes showed that both anti-kinesins bound a 117K band. The migration positions of  $M_r$  standards are indicated at left.

**METHODS.** Production and characterization of the anti-kinesin serum have been described<sup>5</sup>. Microtubule gliding motility was assayed essentially as described<sup>10,14</sup>. Microtubules were polymerized from purified bovine brain tubulin using taxol and stored under liquid N<sub>2</sub> until needed. Coverslips (22 × 22 mm) were treated with 4 µl of chick brain kinesin (prepared as in ref. 5) for 30 min at 20 °C, and then washed with PMED (50 mM PIPES buffer (pH 6.9), 2.5 mM MgSO<sub>4</sub>, 1 mM EGTA, 1 mM DTT). For inhibition studies, 4 µl of IgG diluted in PMED was applied to the coverslip at this point, and after 20 min the coverslip was washed and 4 µl of 50 µg ml<sup>-1</sup> microtubules containing 5 mM magnesium ATP was added. The coverslip was then inverted over a second coverslip and microtubule gliding was observed directly using darkfield optics. The percentage of microtubules exhibiting gliding was determined by observing microtubules on the kinesin-coated surface for 10 s each and scoring whether gliding was detectable; unperturbed microtubule velocities averaged 0.7 µm s<sup>-1</sup>. Macrophages were collected and grown in culture as previously described<sup>19</sup>. Cells were washed and extracted for 3 min using warm 20 mM HEPES buffer (pH 8.0), 250 mM sucrose, 1 mM EGTA, 0.5 mM EDTA, 1 mM DTT, protease inhibitors<sup>20</sup> and 0.02% saponin or 2% SDS; both extracts gave identical immunoblots. Protein was concentrated from the supernatant by chloroform precipitation<sup>21</sup>, electrophoresed, transferred to nitrocellulose, and immunostained as previously described<sup>5</sup>. After immunostaining and photography, blots were stained for total protein using 0.1% India ink in PBS.

There was a clear difference between cells loaded with anti-kinesin and unloaded cells. Cells loaded with polyclonal or monoclonal anti-kinesin antibodies showed a pronounced centripetal collapse and rounding of their tubular lysosomes to form a perinuclear mass, with peripheral regions of the cytoplasm essentially devoid of lysosomes (Fig. 2). The loaded anti-kinesin was largely restricted to the region of the lysosomes, although some was detected near the ruffling edges of the cells (Fig. 2*b*). Neither preimmune IgG loading nor loading with fluorescein-dextran alone produced any apparent collapse of the tubular lysosomes. In addition, anti-kinesin loading had no

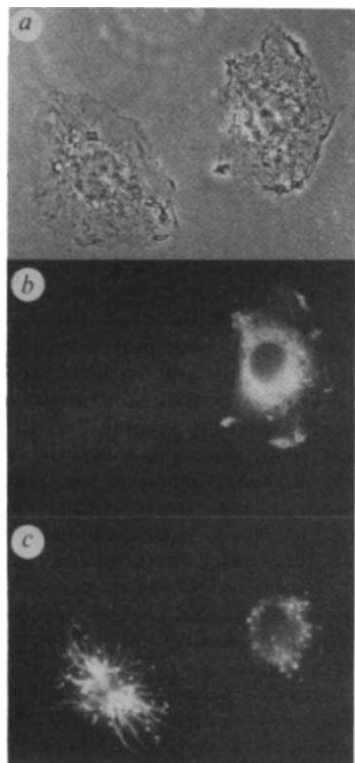


FIG. 2 The effect on the tubular lysosomes of scrape-loading anti-kinesin immunoglobulin into macrophages. The same field is shown here viewed in three ways. A phase contrast micrograph (a) shows two typical cells from a population that had had their tubular lysosomes labelled by uptake of Texas Red-ovalbumin, and had then been scrape-loaded with polyclonal anti-kinesin IgG, replated, fixed, and stained with fluorescein-labelled anti-rabbit immunoglobulin. Epifluorescent illumination in the fluorescein channel (b) revealed which cells had been loaded with anti-kinesin; those which were loaded showed strong fluorescence, as seen in the cell on the right. The anti-kinesin signal was brightest in the perinuclear region and at the ruffling edges of the cells. Texas Red epifluorescence (c) showed an unperturbed array of tubular lysosomes in the unloaded cells, as seen on the left. The anti-kinesin-loaded cells, as seen at right showed a centripetal collapse and rounding up of the lysosomes. Magnification, 1,000 $\times$ .

**METHODS.** IgG was purified to homogeneity from anti-kinesin serum by ammonium sulphate precipitation followed by DEAE Affigel Blue chromatography<sup>22</sup>, and concentrated to 10–15 mg ml<sup>-1</sup>. SUK4 monoclonal IgG was purified from ascitic fluid by protein G-Sepharose chromatography and concentrated to 10 mg ml<sup>-1</sup>. Both antibodies were dialysed against PBS and stored at -80 °C. Macrophages were incubated in 35-mm dishes in medium containing 40  $\mu$ g ml<sup>-1</sup> Texas Red-conjugated ovalbumin for 90 min to label the tubular lysosomes, then washed with unlabelled medium for 30 min. They were then scrape-loaded by first washing with warm PBS, then replacing it with 45–60  $\mu$ l of immunoglobulin in PBS, and scraping the cells off the dish with a rubber policeman<sup>16</sup>. Cells were then replated on coverslips, allowed to reattach and spread for 60 min, and then fixed with 0.05% glutaraldehyde plus 3.7% formaldehyde in PBS. They were then stained with fluorescein isothiocyanate conjugated anti-rabbit immunoglobulin and mounted for viewing and photomicrography.

TABLE 1 Indices of extension of tubular lysosomes after different treatments

| Treatment                                                 | Index of extension<br>$\pm$ s.e.m. (n) | % Extension |
|-----------------------------------------------------------|----------------------------------------|-------------|
| Untreated                                                 | 0.590 $\pm$ 0.020 (26)                 | 100         |
| Fluorescent dextran-loaded                                | 0.606 $\pm$ 0.011 (160)                | 105         |
| Preimmune IgG-loaded<br>(10 mg ml <sup>-1</sup> )         | 0.590 $\pm$ 0.017 (46)                 | 100         |
| Polyclonal immune<br>IgG-loaded (10 mg ml <sup>-1</sup> ) | 0.444 $\pm$ 0.018 (90)                 | 57          |
| SUK4 IgG-loaded<br>(2.5 mg ml <sup>-1</sup> )             | 0.319 $\pm$ 0.018 (40)                 | 20          |
| SUK4 IgG-loaded<br>(10 mg ml <sup>-1</sup> )              | 0.315 $\pm$ 0.017 (39)                 | 18          |
| Nocodazole-treated                                        | 0.253 $\pm$ 0.017 (31)                 | 0           |

The effects of different scrape-loading treatments on the extension of tubular lysosomes were compared quantitatively. By the two sample *t*-test, the indices of extension for fluorescent dextran-loaded and preimmune IgG-loaded cells do not differ from those of untreated controls ( $P > 0.1$ ), whereas the indices for anti-kinesin-loaded cells are significantly reduced ( $P < 0.001$ ) relative to the index for untreated cells. Macrophage tubular lysosomes were labelled with Texas Red-conjugated ovalbumin as described in the Fig. 2 legend. Cells were then scrape-loaded in buffer containing fluorescein-conjugated dextran with an  $M_r$  of 70K. Between 30 and 75 min after replating, cells were observed briefly using epifluorescence in the fluorescein channel to identify loaded cells. These were then photographed in the Texas Red channel to record the distribution of tubular lysosomes, and using phase contrast optics to determine the cell area. Untreated and nocodazole-treated cells were similarly photographed live. The index of extension and percentage of total extension were determined as described in the text. Populations were coded so that all measurements were made without knowledge of which experimental treatment was being analysed.

significant effect on either the degree of cell spreading, or on the apparent density and arrangement of cytoplasmic microtubules (not shown).

To compare the different loading treatments quantitatively, we derived a simple measure of the degree of lysosomal extension for individual cells. For each cell, we measured the area of the polygon formed by circumscribing the lysosomal array and divided it by the total area of the cell on the substratum. This quotient, which we refer to as the 'index of extension', indicates the fraction of the cell's area penetrated by tubular lysosomes. To determine the range of possible values for this index, we compared two populations of macrophages: unperturbed cells, and cells in which the microtubules had been depolymerized by nocodazole, which results in complete collapse of the tubular lysosomes. The index of extension for untreated cells represents a maximum value, whereas that of nocodazole-treated cells represents the minimum possible value (additional centripetal movement of the lysosomal array is prevented by the presence of the nucleus). The range of possible indices of extension runs from 0.253 for completely collapsed lysosomal arrays to 0.590 for unperturbed lysosomes (Table 1).

We then compared populations of cells subjected to each scrape-loading treatment by making blind measurements of their indices of extension, and confirmed our qualitative observations: neither pre-immune immunoglobulin nor fluorescent dextran produced any change in the distribution of tubular lysosomes when loaded into cells, whereas both anti-kinesins produced a significant reduction in the extension of the array (Table 1). The polyclonal anti-kinesin IgG caused a 43% reduction in the extension of the tubular lysosomes, whereas 2.5 mg ml<sup>-1</sup> and 10 mg ml<sup>-1</sup> SUK4 IgG caused 80% and 82% reductions in extension, respectively.

The scrape-loading procedure resuspends cells as it introduces macromolecules into their cytoplasm, so our subsequent assay of radial lysosome extension actually measures the re-establishment of the lysosomal array when cells are replated. Thus,



kinesin could serve primarily to establish that array but not to maintain it. The behaviour of lysosomes observed in living cells indicates, however, that the lysosome distribution is maintained by a continuous balance of outward and inward vesicle movements. We therefore suggest that kinesin ATPase has a role in the formation and maintenance of the lysosomal compartment in the form of an extended tubulovesicular array in macrophages. On the basis of immunolocalization<sup>5,17</sup> and *in vitro* motility studies<sup>3,4,13,18</sup> this probably involves kinesin molecules anchored to the organelle surface which undergo ATP-dependent interaction with microtubules to produce centrifugally directed translocating forces. □

Received 15 January; accepted 15 June 1990.

- Swanson, J. A., Bushnell, A. & Silverstein, S. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1921–1925 (1987).
- Lee, C. & Chen, L. B. *Cell* **54**, 37–46 (1988).
- Dabora, S. L. & Sheetz, M. P. *Cell* **54**, 27–35 (1988).
- Vale, R. D. & Hotani, H. *J. Cell Biol.* **107**, 2233–2241 (1988).
- Hollenbeck, P. J. *J. Cell Biol.* **108**, 2335–2342 (1989).
- Sheetz, M. P. *Bioessays* **7**, 165–168 (1987).
- Vale, R. D. *A. Rev. Cell Biol.* **3**, 347–378 (1987).
- Herman, B. & Albertini, D. F. *J. Cell Biol.* **98**, 565–576 (1984).
- Matteoni, R. & Kreis, T. E. *J. Cell Biol.* **105**, 1253–1265 (1987).
- Vale, R. D., Reese, T. S. & Sheetz, M. P. *Cell* **42**, 39–50 (1985).
- Brady, S. T. *Nature* **317**, 73–75 (1985).
- Hollenbeck, P. J. *Protoplasma* **145**, 145–152 (1988).
- Vale, R. D. *et al. Cell* **43**, 623–632 (1985).
- Ingold, A. L., Cohn, S. A. & Scholey, J. M. *J. Cell Biol.* **107**, 2657–2667 (1988).
- Scholey, J. M., Heuser, J., Yang, J. T. & Goldstein, L. S. B. *Nature* **338**, 355–357 (1989).
- McNeil, P. L., Murphy, R. F., Lanni, F. & Taylor, D. L. *J. Cell Biol.* **98**, 1556–1564 (1984).
- Pfister, K. K., Wagner, M. C., Stenoi, D. L., Brady, S. T. & Bloom, G. S. *J. Cell Biol.* **108**, 1453–1463 (1989).
- Schroer, T. A., Schnapp, B. J., Reese, T. S. & Sheetz, M. P. *J. Cell Biol.* **107**, 1785–1792 (1988).
- Swanson, J. A. *J. Cell Sci.* **94**, 135–142 (1989).
- Hollenbeck, P. J., Suprynowicz, F. & Cande, W. Z. *J. Cell Biol.* **99**, 1251–1258 (1984).
- Wessel, D. & Flügge, U. I. *Analyst Biochem.* **138**, 141–143 (1984).
- Bruck, C., Portetelle, D., Glineur, C. & Bollen, A. *J. Immunol. Meth.* **53**, 313–319 (1982).

ACKNOWLEDGEMENTS. We thank H. McNamee for assistance with preliminary experiments, J. M. Scholey and M. P. Sheetz for advice on *in vitro* motility assays, and J. M. Scholey for the gift of SUK4 hybridoma cells. This work was supported by the NIH.

## Telomere reduction in human colorectal carcinoma and with ageing

Nicholas D. Hastie, Maureen Dempster,  
Malcolm G. Dunlop\*, Alastair M. Thompson\*,  
Daryll K. Green & Robin C. Allshire\*

MRC Human Genetics Unit, Western General Hospital, Edinburgh, UK

WE have hypothesized that end-to-end chromosome fusions observed in some tumours could play a part in genetic instability associated with tumorigenesis and that fusion may result from the loss of the long stretches of G-rich repeats found at the ends of all linear chromosomes<sup>1</sup>. We therefore asked whether there is telomere loss or reduction in common tumours. Here we show that in most of the colorectal carcinomas that we analysed, there is a reduction in the length of telomere repeat arrays relative to the normal colonic mucosa from the same patient. We speculate on the consequences of this loss for tumorigenesis. We also show that the telomere arrays are much smaller in colonic mucosa and blood than in fetal tissue and sperm, and that there is a reduction in average telomere length with age in blood and colon mucosa. We propose that the telomerase<sup>2–4</sup> is inactive in somatic tissues, and that telomere length is an indicator of the number of cell divisions that it has taken to form a particular tissue and possibly to generate tumours.

\* Present addresses: Department of Surgery, Royal Infirmary, Edinburgh EH3 9YW, UK (M.G.D. and A.M.T.), and Cold Spring Harbor Laboratories, Cold Spring Harbor, New York 11724, USA (R.C.A.).

The termini of human chromosomes, like those of other eukaryotes, comprise stretches of G-rich repeats<sup>5–10</sup> which are of ~10 kilobases (kb) and 15 kb in blood and sperm respectively<sup>5,7,10,11</sup>. When DNA is digested with a frequently cutting restriction enzyme such as *HinfI* or *AluI*, which will not cut within the repeat, subsequent probing with a labelled oligonucleotide for (TTAGGG)<sub>n</sub> will reveal the size of the terminal repeat array (TRA). Figure 1a shows the hybridization of a (TTAGGG)<sub>4</sub> probe to *HinfI*-digested DNA from matched normal colonic mucosa and colon carcinoma samples from seven individuals. There is some variation in the average length of the TRA between normal mucosae from different individuals but not much difference in signal strength. In all the carcinomas there is a reduction in length of the repeat stretch relative to the respective normal mucosa and in some cases there is a marked reduction in the signal intensity.

We repeated the experiment with the restriction enzyme *AluI* and obtained a very similar result to that obtained with *HinfI*

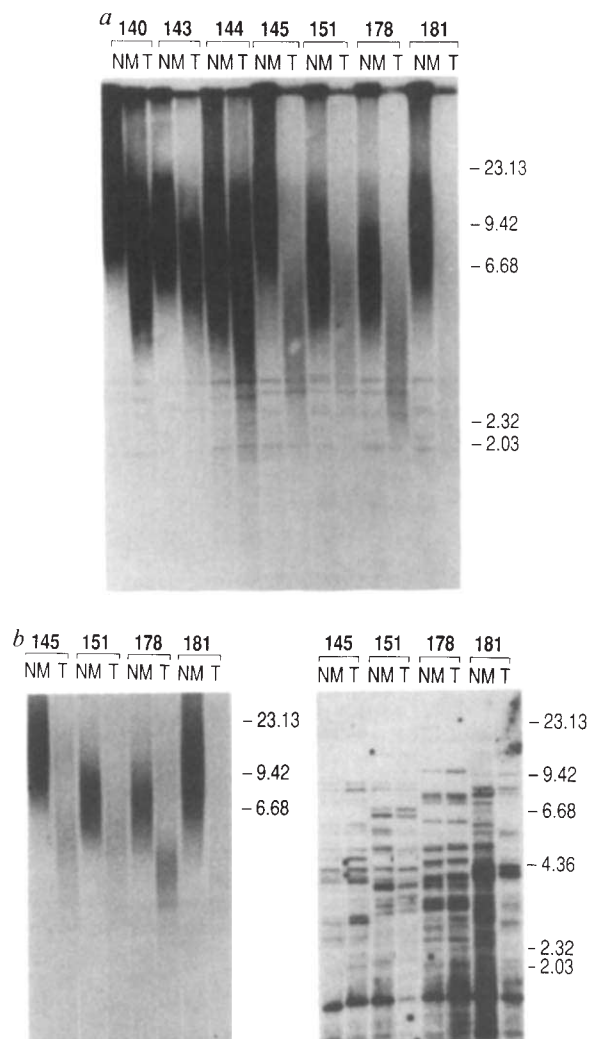


FIG. 1 Southern blot analysis of telomere repeat arrays in colorectal carcinomas and matched blood DNA samples. *a*, DNA was digested with the restriction enzyme *HinfI* and hybridized with <sup>32</sup>P-labelled (TTAGGG)<sub>4</sub> probe. *b*, DNA was digested with the restriction endonuclease *AluI* and hybridized with a <sup>32</sup>P-labelled (TTAGGG)<sub>4</sub> probe (left panel) or a (CAC)<sub>5</sub> probe (right panel). The filter was first hybridized to (CAC)<sub>5</sub> then stripped and rehybridized to (TTAGGG)<sub>4</sub>. NM, normal mucosa; T, tumour; numbers refer to patient numbers. Size markers are indicated to the right (kb).

METHODS. Either *HinfI*- or *AluI*-digested total genomic human DNA (5 µg) was loaded onto a 0.8% agarose gel. After electrophoresis the gel was blotted to Hybond and hybridized to the radiolabelled probes. Hybridization was at 48 °C in 5 × SSC buffer and washes were at 48 °C in 4 × SSC.

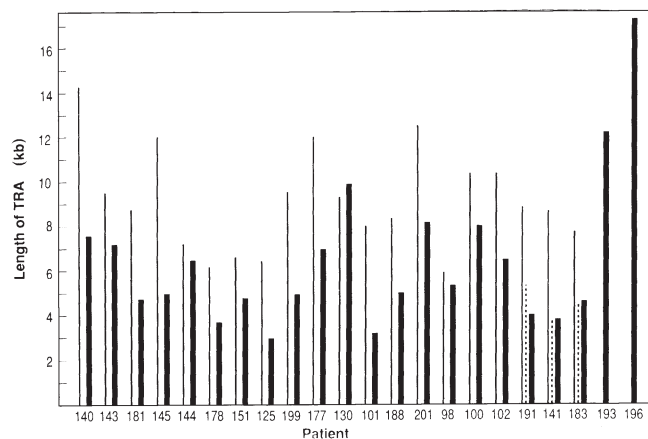


FIG. 2 Mean size estimates for the TRA in colorectal carcinoma, matched normal mucosae and adenomas. —, Normal mucosa; ■, carcinoma; ---, adenoma.

**METHODS.** Autoradiographs such as those shown in Fig. 1 were analysed by an automatic autoradiographic scanner as described previously<sup>7</sup>. In all cases *Hinf*I-digested DNAs were analysed but similar results were obtained with *Alu*I. The optical density reading under the peak was integrated and the mean determined. Usually a gaussian distribution was obtained and the mean position was very close to that at the centre of the peak. In all cases the values obtained for several different experiments were within 5–10% of each other. To avoid problems of over exposure of the signal, various exposure times were analysed. In some cases two separate peaks were observed in the tumour samples (for example, 145), the larger peak, of lower optical density, similar in size to the corresponding normal mucosa. We have attributed this to contamination of the tumour by normal mucosa, for which there is independent evidence from allele loss studies. In these cases the mean size of the smaller peak is given.

(Fig. 1b). DNA loading and integrity were checked by prehybridizing all blots to a (CAC)<sub>5</sub> minisatellite probe (Fig. 1b, right panel).

To provide an estimate of the degree of reduction of the TRA in tumours, densitometric scans were carried out on these autoradiographs (Fig. 1) and those obtained for 13 other tumours and matched normal mucosae. The average TRA length in the 20 normal mucosae and matching carcinomas is depicted in Fig. 2; in all but one tumour (130) there was a reduction in the average TRA length relative to the normal mucosa.

In three patients we were able to study the length of the TRA in adenomas, which are the benign precursors of carcinomas. In all three cases (191, 141, 183) there was a reduction in the adenoma similar in extent to that found in the corresponding carcinoma. In two carcinomas (193, 196) out of 35 studied, the telomeres were very much larger. Unfortunately, no matched normal mucosae were available for these samples so we could not determine if this represented an increase in telomere length in the tumour.

In several cases, the average TRA length is close to 3 kb, with molecules ranging to 1 kb or less. After correcting for DNA loading using internal controls, there was a reduction in signal intensity of about 90%, 90%, 80%, 65%, 65% and 20% for tumours 181, 145, 178, 151, 143 and 144, respectively. Thus it is reasonable to suppose that in a proportion of cells in tumours and adenomas (Fig. 2) there will be one or several chromosome ends lacking G-rich repeats.

In a preliminary series of experiments we have observed TRA reduction in ovarian and lung carcinomas as well as the colon carcinomas reported here. Telomere shortening has also been reported in two Wilms' tumours<sup>10</sup>.

What is the mechanism underlying this TRA reduction? Reduction could reflect the number of cell divisions it has taken to produce the tumour cells in the absence of telomerase func-

tion<sup>2–4</sup>, which may be restricted to germ cells. We have compared the average TRA length in sperm with that in tissues from a 26-week human fetus and colon mucosa (Fig. 3) and blood (data not shown) samples analysed in this study. Although tissue from only a single fetus was studied, it is clear that TRA length in these tissues is close (1–2-kb shorter) to that found in sperm and considerably larger than that found in any colon mucosa (on average 12 kb shorter than sperm) or blood (on average 14 kb shorter) sample.

Firm figures are not available, but it is likely that the tissues of a developed fetus result from 20–50 cell divisions, whereas several hundred or thousands of divisions have produced the colonic mucosa and blood cells of 60-year old individuals<sup>12</sup>. Thus the degree of telomere reduction is more or less proportional to the number of cell divisions. It has been shown that the ends of *Drosophila* chromosomes without normal telomeres reduce in size by ~4 base pairs (bp) per cell division<sup>13</sup> and that the ends of yeast chromosomes reduce by a similar degree in a mutant presumed to lack telomerase function<sup>14</sup>. If we assume the same rate of reduction is occurring during somatic division in human tissues, then a reduction in TRA by 14 kb would mean that 3,500 ancestral cell divisions lead to the production of cells in the blood of a 60-year old individual; using estimates of sperm telomere length found elsewhere<sup>10</sup> we obtain a value of 1,000–2,000. These values compare favourably with those postulated for mouse blood cells<sup>12</sup>. Thus, we propose that telomerase is indeed lacking in somatic tissues. In this regard it is of interest to note that in maize, broken chromosomes are only healed in sporophytic (zygotic) tissues and not in endosperm (terminally differentiated)<sup>15</sup>, suggesting that telomerase activity is lacking in the differentiated tissues.

Consistent with this idea, there is a significant reduction of TRA length as a function of age in human blood DNA, although there is a high degree of scatter (Fig. 4). From the regression analysis we can calculate that the rate of reduction is 33 bp per year. Extending this idea to colon tumours, a reduction in 5 kb would mean that it has taken ~1,000 cell divisions to produce a tumour from normal mucosa. Analysis of three cases (Fig. 2; 191, 141, 183) suggests that most of these doublings would take place in the formation of the adenoma. In tumours 193 and 196 we assume that the telomerase has been reactivated as proposed

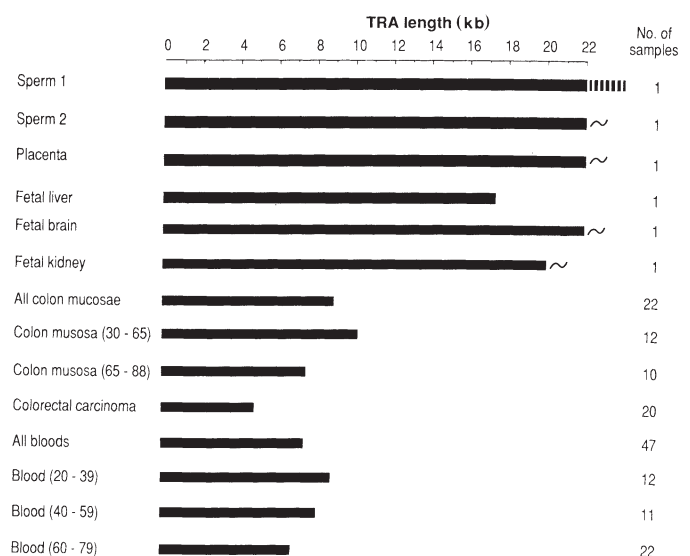


FIG. 3 Telomere length in different tissues and in association with ageing. The estimates are averages for each tissue where more than one sample was analysed. The data was obtained as described in Figs 1 and 2. Vertical bars indicate greater than, and the symbol ~ means approximately (values of ~20 kb and greater are approximations). Numbers in brackets in the left column refer to age range.



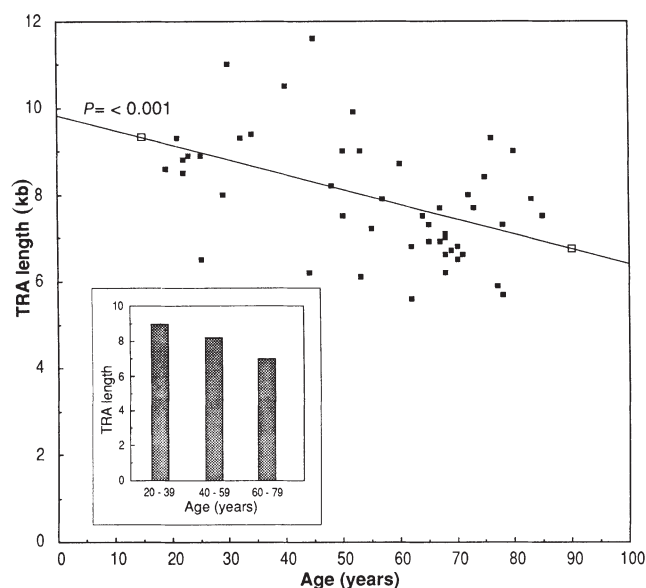


FIG. 4 Telomere reduction with age in human blood. The straight line was obtained by regression analysis. The significance of the regression,  $t = -3.63$ ;  $P < 0.001$ . The slope is equivalent to 33 bp per year.

for HeLa cells in culture, which are known to contain telomerase activity<sup>4,10</sup>.

One alternative explanation for our observations is that in tumours the cells with shorter telomeres have a growth advantage over those with larger telomeres, a situation described for vegetative cells of tetrahymena<sup>16</sup>.

Telomere losses could be a consequence of tumorigenesis or be involved in causing malignancy. We believe there are two ways by which TRA losses could contribute to genetic instability, which is important for tumorigenesis. First, chromosomes lacking terminal repeats may be less stable and lost at higher frequency, a situation for which there is a precedent in the yeast EST1 mutant<sup>14</sup>. Second, chromosomes without telomeres may be prone to fusion-bridge-breakage cycles as described by McClintock<sup>17</sup>, leading to daughter cells which receive partly deleted and partly duplicated chromosomes, respectively. There is no direct evidence to support this idea in carcinogenesis, although in senescent cells where chromosome fusions are found at high frequency<sup>18</sup> a loss of telomere repeats has been reported<sup>19</sup>. Also, telomere fusions occur in a range of tumours, including histiocytomas<sup>20</sup>, cardiac myxomas<sup>21</sup>, renal tumours<sup>22</sup> and leukaemias<sup>23</sup>, as well as the common cancers at much lower frequency. In addition, broken ends within amplified dihydrofolate reductase genes lacking telomeres will fuse to other chromosomes, setting up breakage-fusion-bridge cycles<sup>24</sup>.

Deletions or losses by these routes could play a part in generating the loss of alleles of restriction fragment length polymorphisms that often occurs in colorectal carcinomas and that is thought to reflect the need to render mutations in tumour suppressor genes homozygous<sup>25,26</sup>. Although some of these allele losses occur in the adenoma stage the majority are not apparent until the carcinoma stage<sup>25</sup>. In the three cases studied here (191, 141, 183; Fig. 2) we find that telomere reduction has already taken place by the adenoma stage. Hence these losses could be involved in causing the chromosome rearrangements subsequently observed in the carcinoma. □

1. Allshire, R. C. *et al.* *Nature* **332**, 656-659 (1988).
2. Moyzis, R. K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 6222-6226 (1988).
3. Allshire, R. C., Dempster, M. & Hastie, N. D. *Nucleic Acids Res.* **17**, 4611-4627 (1989).
4. Brown, W. R. A. *Nature* **338**, 774-776 (1989).
5. Cross, S. H., Allshire, R. C., McKay, S., McGill, N. & Cooke, H. J. *Nature* **338**, 771-774 (1989).
6. de Lange, T. *et al.* *Molec. cell Biol.* **10**, 518-527 (1990).
7. Cooke, H. J. & Smith, B. A. *Cold Spring Harb. Symp. quant. Biol.* **6**, 213-219 (1986).
8. Potten, C. S. & Morris, R. J. *J. Cell Sci. Suppl.* **10**, 45-62 (1988).
9. Levis, R. W. *Cell* **58**, 791-801 (1989).
10. Lundblad, V. & Szostak, J. W. *Cell* **57**, 633-643 (1989).
11. McClintock, B. *Genetics* **26**, 234-282 (1941).
12. Larson, D. D., Spangler, E. A. & Blackburn, E. H. *Cell* **50**, 477-483 (1987).
13. McClintock, B. *Cold Spring Harb. Symp. Quant. Biol.* **16**, 13-49 (1951).
14. Bann, P. A. *Am. J. Hum. Genet.* **28**, 465-473 (1976).
15. Harley, C. B., Futcher, A. B. & Greider, C. W. *Nature* **345**, 458-460 (1990).
16. Mandahl, N. *et al.* *Hum. Genet.* **71**, 321-324 (1985).
17. Dewald, C. W. *et al.* *Mayo Clin. Proc.* **62**, 558-567 (1987).
18. Kovacs, G., Muller-Brechlin, R. & Szucs, S. *Cancer Genet. Cytogenet.* **28**, 363-365 (1987).
19. Morgan, R. *et al.* *Hum. Genet.* **73**, 260-263 (1986).
20. Kaufman, R. J., Sharp, P. A. & Latt, S. A. *Molec. cell Biol.* **3**, 699-711 (1983).
21. Vogelstein, B. *et al.* *New Engl. J. Med.* **319**, 525-532 (1988).
22. Vogelstein, B. *et al.* *Science* **244**, 207-211 (1989).

ACKNOWLEDGEMENTS. We thank H. Cooke for fetal DNA samples and for helpful discussions, C. Potten and J. Ansell for discussions about stem-cell turnover, A. Carothers for statistical analysis, C. Harley, B. Futcher and C. Greider for communicating their work before publication, Professor H. J. Evans and Dr M. Steel and their support and discussions, D. Stewart and N. Davidson for photography and graphics and L. Campbell and E. Kay for secretarial assistance.

## A truncated human chromosome 16 associated with $\alpha$ thalassaemia is stabilized by addition of telomeric repeat (TTAGGG)<sub>n</sub>

Andrew O. M. Wilkie\*, Janette Lamb\*, Peter C. Harris\*, Roger D. Finney† & Douglas R. Higgs\*

\* MRC Molecular Haematology Unit, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

† Department of Haematology, North Tees General Hospital, Stockton-on-Tees, Cleveland TS19 8PE, UK

THE instability of chromosomes with breaks induced by X-irradiation led to the proposal that the natural ends of chromosomes are capped by a specialized structure, the telomere<sup>1</sup>. Telomeres prevent end-to-end fusions and exonucleolytic degradation, enable the end of the linear DNA molecule to replicate, and function in cell division (reviewed in ref. 2). Human telomeric DNA comprises ~2-20 kilobases (kb) of the tandemly repeated sequence (TTAGGG)<sub>n</sub>, oriented 5' → 3' towards the end of the chromosome<sup>3,4</sup>, interspersed with variant repeats in the proximal region<sup>5</sup>. Immediately subtelomeric lie families of unrelated repeat motifs (telomere-associated sequences) whose function, if any, is unknown<sup>6,7</sup>. In lower eukaryotes the formation and maintenance of telomeres may be mediated enzymatically (by telomerase)<sup>8</sup> or by recombination<sup>9</sup>; in man the mechanisms are poorly understood, although telomerase has been identified in HeLa cells<sup>4</sup>. Here we describe an  $\alpha$  thalassaemia<sup>10</sup> mutation associated with terminal truncation of the short arm of chromosome 16 (within band 16p13.3) to a site 50 kb distal to the  $\alpha$  globin genes, and show that (TTAGGG)<sub>n</sub> has been added directly to the site of the break. The mutation is stably inherited, proving that telomeric DNA alone is sufficient to stabilize the broken chromosome end. This mechanism may occur in any genetic disease associated with chromosome truncation.

During a search for novel  $\alpha$  thalassaemia determinants, we identified a British man (II-1) with haemoglobin H (HbH) disease (HbH level 5.3%). This relatively severe form of  $\alpha$  thalassaemia is usually caused by deletion of three of the normal

Received 12 April; accepted 14 June 1990.

1. Hastie, N. D. & Allshire, R. C. *Trends Genet.* **5**, 326-331 (1989).
2. Greider, C. W. & Blackburn, E. H. *Cell* **43**, 405-413 (1985).
3. Greider, C. W. & Blackburn, E. H. *Nature* **337**, 331-337 (1989).
4. Morin, G. B. *Cell* **59**, 521-529 (1989).

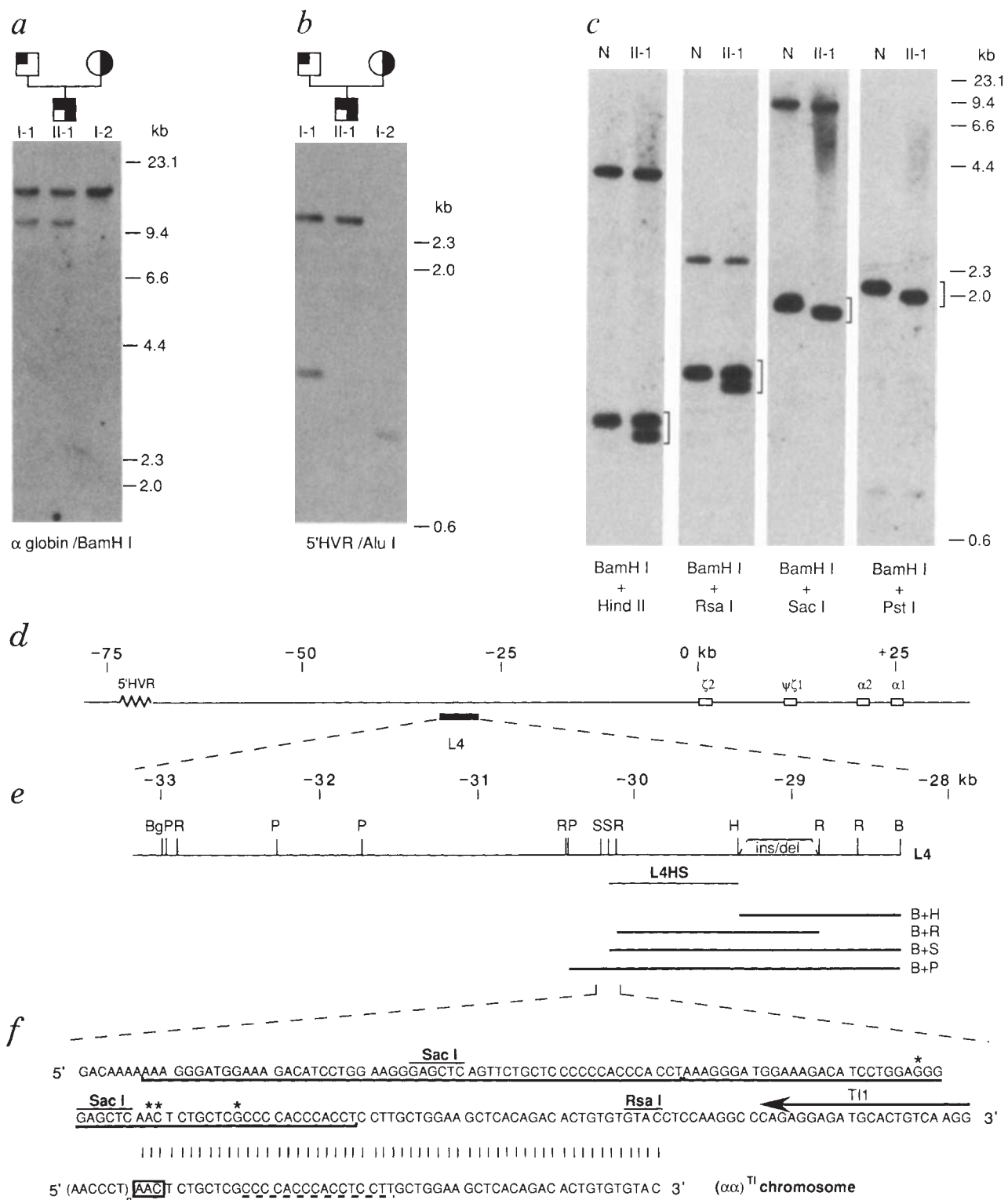
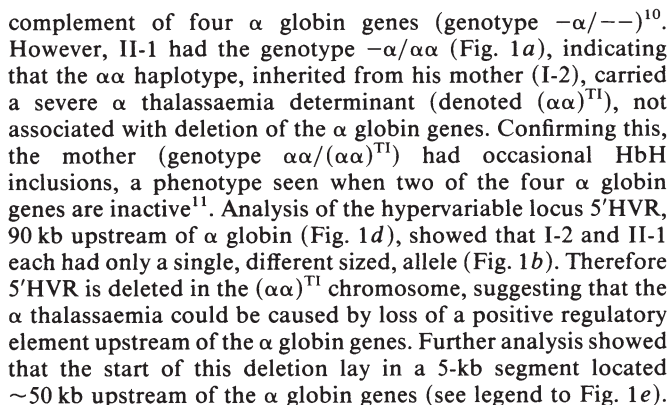


FIG. 1 Localization of breakpoint in the mutant thalassaemia chromosome ( $\alpha\alpha$ )<sup>Tl</sup>. **a**, Inheritance of  $\alpha$  globin alleles. The 14-kb and 10.5-kb bands on a BamHI digest are diagnostic of  $\alpha\alpha$  and  $-\alpha$  chromosomes, respectively<sup>10</sup>. **b**, Inheritance of  $\alpha$  globin 5'HVR (ref. 31) alleles. **c**, Inheritance of L4 (ref. 32) in II-1 compared with a normal control (N) homozygous for the larger allele of the insertion-deletion polymorphism. Square brackets denote fragments corresponding to this polymorphism. **d**, Map of  $\alpha$  globin region, showing positions of  $\alpha$  and  $\zeta$ -like globin genes, 5'HVR and L4. 0 kb corresponds to the position of the  $\zeta$ 2 messengerRNA cap site. **e**, Map of L4 region, showing the site of the insertion-deletion polymorphism (ins-del) (frequency of larger allele 0.97;  $N=98$  chromosomes), the position of probe L4HS, and the polymorphic restriction fragments marked in Fig. 1c (an internal RsaI fragment is detected on the BamHI-RsaI double digest). The breakpoint was initially localized to L4 by DNA dosage, which detected both copies of the

BamHI site at coordinate -28.3, but only a single copy of the BglII site at -33.0. Abbreviations: B, BamHI; Bg, BglII; H, HindII; P, PstI; R, RsaI; S, SacI. **f**, Sequence across breakpoint of ( $\alpha\alpha$ )<sup>Tl</sup> chromosome (bottom line) compared with the normal sequence (upper two lines). The normal sequence contains a pair of 56-bp tandem repeats (underlined); the four mismatches between these repeats are marked by asterisks against the 3' repeat. Primer Tl1 is arrowed. The ( $\alpha\alpha$ )<sup>Tl</sup> sequence contains a 3-bp ambiguity in the start of telomere addition (boxed); this coincides with two of the tandem-repeat mismatches. The internal telomere-like sequence is denoted by a dashed line. A total of 470 bp of flanking sequence showed no other features of note. METHODS. Southern blotting and hybridization were carried out as previously described<sup>33</sup>. Probe L4 was hybridized in phosphate buffer<sup>34</sup> with 400  $\mu$ g ml<sup>-1</sup> sonicated human DNA<sup>35</sup>.





To analyse the sequence at the breakpoint of the  $(\alpha\alpha)^{\text{TI}}$  chromosome, we amplified the DNA across the breakpoint (see legend to Fig. 3). Southern hybridization with L4HS showed a smear of specific product (size 0.1–3 kb) in samples from I-2 and II-1, but not in a normal control (not illustrated). Sequencing of cloned amplification products from I-2 and II-1 showed

identical breakpoints, with canonical telomeric sequence (TTAGGG)<sub>n</sub> starting immediately beyond the break (Fig. 3; patient I-2).

AGGGTTCCTCCGAGCTCGCGAGC

**METHODS.** The breakpoint was amplified by the polymerase chain reaction using the primers T11 (5'-CCTTGCACAGTGCATCTCCTCTG-3'), located within L4 (see Fig. 1f) and TEL1 (5'-TATGGATCCCTCAACCCGACCCCTAACCC-3'), which incorporates a single A → G mismatch (G) in the canonical C-strand telomere sequence and a *Bam*HI site (underlined) near the 5' end. Amplification in a Perkin Elmer Cetus DNA Thermal Cycler used Amplitaq (Cetus) in 3 mM Mg<sup>2+</sup> and 10% dimethylsulphoxide, with an annealing temperature of 58 °C for two cycles followed by 65 °C for 29 cycles. The product was digested with *Bam*HI and *Rsa*I and cloned into *Bam*HI-*Hind*III-cut pUC12 and pUC13. Positive clones were selected by colony screening using fragment L4HS (ref. 33). The corresponding normal sequence was obtained by cloning the *Rsa*I and *Hind*III-*Sac*I fragments which overlap the breakpoint. Double-stranded<sup>36</sup> and M13 sequencing were performed using the Sequenase kit (USB). Three clones from I-2 and one from II-1 were sequenced; all showed an identical breakpoint, and the total of 43 telomeric repeats within the clones (excluding the TEL1 primer) were all of the canonical type. The mismatch in the TEL1 primer serves both as an anchor during amplification and an identification tag for the primer, and thus excludes the possibility that the TTAGGG repeats have arisen artefactually from the primer sequence.

have described may be a relatively frequent cause of human genetic disease: for instance, analogous deletions of chromosomes 4p and 17p may account for some cases of the Wolf-Hirschhorn<sup>13</sup> and Miller-Dieker<sup>14</sup> syndromes, respectively. The harmful effects of such truncations may result from loss of both gene function and chromosome stability. The viability of the  $(\alpha\alpha)^{TT}$  mutation doubtlessly relates to the close proximity (within 400 kb) of the  $\alpha$  globin complex to the telomere<sup>15</sup>. The absence of clinical abnormalities (apart from  $\alpha$  thalassaemia) contrasts with the effects of more extensive deletions of 16p13.3 (ref. 16), implying that monosomy for the region distal to the  $\alpha$  globin complex is otherwise of little phenotypic consequence. The  $\alpha$  thalassaemia probably results from loss of a positive regulatory region upstream of the  $\alpha$  globin complex, analogous to that described for the  $\beta$  globin cluster<sup>17</sup>. This is supported both experimentally<sup>18</sup> and by the existence of an interstitial deletion of overlapping upstream sequences that causes  $\alpha$  thalassaemia<sup>19</sup>. The  $(\alpha\alpha)^{TT}$  mutation establishes the orientation of the  $\alpha$  globin complex with respect to chromosome 16, with the 5' end distal: a result confirmed by pulsed-field mapping (P.C.H., unpublished data).

The stabilization of broken chromosomes has been most fully studied in the ciliated protozoa, in which this is an integral feature of normal development: telomeric sequence is added *de novo* to the ends of macronuclear DNA in *Paramecium*<sup>20,21</sup> and *Tetrahymena*<sup>22</sup>. The relevance of these findings to the problem of sporadic DNA breakage and repair in man is uncertain. Terminal deletions more analogous to that characterized here have been described in *Drosophila* and *Plasmodium*, and four different structural bases have been found. In *Drosophila*, sometimes no repair of the broken end occurs and there is loss of ~70 bp of terminal sequence per generation<sup>23,24</sup>. On other occasions, addition of DNA belonging to the HeT family is observed<sup>25</sup>. In *Plasmodium*, stabilization by conventional telomeric sequence has been reported: this may be mediated either by recombination between subtelomeric repeats<sup>26</sup>, or by *de novo* addition of telomere sequence<sup>27,28</sup>. Our findings resemble the last of these mechanisms; the immediate transition to telomeric sequence and the stable inheritance of the mutation through one generation indicates that telomeric sequence, in the absence of telomere-associated sequence, is alone sufficient to stabilize the chromosome end.

The mechanism of the addition of  $(TTAGGG)_n$  is uncertain. The enzyme activity identified in HeLa cells<sup>4</sup>, like other telomerases<sup>2</sup>, requires a G-rich single-stranded oligonucleotide to prime the addition of telomeric repeat. Although a short 3' overhang could have been generated on the newly broken  $(\alpha\alpha)^{TT}$  chromosome after DNA replication<sup>29</sup>, the sequence just proximal to the site of telomere addition includes two C residues on this strand (that is, G residues in Fig. 1f) and seems unlikely to be a suitable primer. In the ciliates, a similar disparity has been noted between the apparently specific priming requirements of telomerase and the nonspecific nature of the sequences at which new telomeres are actually formed<sup>20-22</sup>. Perhaps human telomerase can recognize internal telomere-like sequences, as proposed for *Saccharomyces*<sup>30</sup>; such a sequence is present ~10-25 bp proximal to the breakpoint of the  $(\alpha\alpha)^{TT}$  chromosome (Fig. 1f). Alternatively a mechanism involving recombination<sup>9</sup> may be involved. Clearly, further analysis of the termini from natural broken chromosomes is needed to establish the sequence requirements, if any, for telomere healing in man. □

8. Yu, G.-L., Bradley, J. D., Attardi, L. D. & Blackburn, E. H. *Nature* **344**, 126-132 (1990).
9. Wang, S.-S. & Zakian, V. A. *Nature* **345**, 456-458 (1990).
10. Higgs, D. R. *et al. Blood* **73**, 1081-1104 (1989).
11. Galanello, R., Paglietti, E., Melis, M. A., Giagu, L. & Cao, A. *Acta Haematol., Basel* **72**, 34-36 (1984).
12. Yao, M.-C. & Yao, C.-H. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7436-7439 (1981).
13. Gusella, J. F. *et al. Nature* **318**, 75-78 (1985).
14. Ledbetter, D. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 5136-5140 (1989).
15. Harris, P. C., Barton, N. J., Higgs, D. R., Reeders, S. T. & Wilkie, A. O. M. *Genomics* **7**, 195-206 (1990).
16. Wilkie, A. O. M. *et al. Am. J. hum. Genet.* **46**, 1112-1126 (1990).
17. Grosveld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. *Cell* **51**, 975-985 (1987).
18. Higgs, D. R. *et al. Genes Dev.* (in the press).
19. Hatton, C. S. R. *et al. Blood* **76**, 221-227 (1990).
20. Baroin, A., Prat, A. & Caron, F. *Nucleic Acids Res.* **15**, 1717-1728 (1987).
21. Forney, J. D. & Blackburn, E. H. *Molec. cell. Biol.* **8**, 251-258 (1988).
22. Spangler, E. A., Ryan, T. & Blackburn, E. H. *Nucleic Acids Res.* **16**, 5569-5585 (1988).
23. Levis, R. W. *Cell* **58**, 791-801 (1989).
24. Biessmann, H., Carter, S. B. & Mason, J. M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1758-1761 (1990).
25. Biessmann, H. *et al. Cell* **61**, 663-673 (1990).
26. Corcoran, L. M., Thompson, J. K., Walliker, D. & Kemp, D. J. *Cell* **53**, 807-813 (1988).
27. Pologe, L. G. & Ravetch, J. V. *Cell* **55**, 869-874 (1988).
28. Cappai, R. *et al. Molec. cell. Biol.* **9**, 3584-3587 (1989).
29. Watson, J. D. *Nature New Biol.* **239**, 197-201 (1972).
30. Murray, A. W., Claus, T. E. & Szostak, J. W. *Molec. cell. Biol.* **8**, 4642-4650 (1988).
31. Jarman, A. P. & Higgs, D. R. *Am. J. hum. Genet.* **43**, 249-256 (1988).
32. Nicholls, R. D., Fischel-Ghodsian, N. & Higgs, D. R. *Cell* **49**, 369-378 (1987).
33. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning* 2nd edn (Cold Spring Harbor Laboratory, New York, 1989).
34. Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991-1995 (1984).
35. Sealey, P. G., Whittaker, P. A. & Southern, E. M. *Nucleic Acids Res.* **13**, 1905-1922 (1985).
36. Chen, E. Y. & Seeburg, P. H. *DNA* **4**, 165-170 (1985).

ACKNOWLEDGEMENTS. We thank the family for their cooperation with this study. H. Ayyub for technical assistance, J. P. Bennett and S. L. Thein for advice on polymerase chain reaction, V. J. Buckle for the cytogenetic analysis, W. R. A. Brown for stimulating criticism, P. Vyas for reading the manuscript, E. Rose for typing, and Sir David Weatherall for continued support. This work was supported by the MRC, Wellcome Trust and Action Research for the Crippled Child.

## A bacterial virulence determinant encoded by lysogenic coliphage $\lambda$

James J. Barondess\*<sup>†</sup> & Jon Beckwith<sup>†</sup>

\* Program in Cell and Developmental Biology and † Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

ALTHOUGH phage  $\lambda$  represents a well studied biological system, it has certain features that remain obscure. Among these is the function of the roughly one third of the phage genome dispensable for growth in the laboratory, yet retained despite undoubted pressure to economize<sup>1</sup>. Here we report that these 'accessory' sequences contain two genes which are expressed during lysogeny, and encode host-cell envelope proteins. One of these is *lom* (ref. 2), the product of which is found in the bacterial outer membrane, and is homologous to virulence proteins of two other enterobacterial genera. The other gene, previously unidentified, we designate *bor*. Expression of *bor* significantly increases the survival of the *Escherichia coli* host cell in animal serum. This property is a well known bacterial virulence determinant<sup>3,4</sup>—indeed, *bor* and its adjacent sequences are highly homologous to the *iss* serum resistance locus of the plasmid ColV2-K94, which confers virulence in animals<sup>5,6</sup>. These results show that the  $\lambda$  prophage is more transcriptionally active than has long been assumed, and suggest that lysogeny may generally have a role in bacterial survival in animal hosts, and perhaps in pathogenesis.

We generated a random collection of fusions to the alkaline phosphatase gene in the transducing phage  $\lambda$ 16-25 (ref. 7) using the transposon *TnphoA*. Such fusions allow the identification of proteins and segments of proteins that contain signals for export beyond the cytoplasm, as the alkaline phosphatase portion of the hybrid molecules is only enzymatically active when exported (for a review of this approach and its uses see ref. 8). Two classes of fusions to  $\lambda$  genes were found. Those with lower *phoA* activity (3-4 units) mapped to *lom* (18,965-19,582 in the  $\lambda$  sequence<sup>9</sup>), whereas those with higher activity (100 units) mapped to the extreme right end of the phage DNA, distal to the *R<sub>z</sub>* gene, and were oriented leftward. This region of  $\lambda$  has

Received 11 May; accepted 12 July 1990.

1. Muller, H. J. *Collect. Net.* **13**, 182-193 (1938).
2. Zakian, V. A. *et al. Rev. Genet.* **23**, 579-604 (1989).
3. Moyzis, R. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 6622-6626 (1988).
4. Morin, G. B. *Cell* **59**, 521-529 (1989).
5. Allshire, R. C., Dempster, M. & Hastie, N. D. *Nucleic Acids Res.* **17**, 4611-4627 (1989).
6. Cheng, J.-F., Smith, C. L. & Cantor, C. R. *Nucleic Acids Res.* **17**, 6109-6127 (1989).
7. de Lange, T. *et al. Molec. cell. Biol.* **10**, 518-527 (1990).



not previously been thought to have any coding function. These fusions thus identify a new gene, which we call *bor* (as in blue open reading frame, as the lysogens are blue on alkaline phosphatase XP (5-bromo-4-chloro-3-indolyl phosphate) indicator plates). Both classes of fusions are expressed in lysogens.

Representative fusions of each class were sequenced (Fig. 1). The predicted *bor* protein product of 97 amino acids has a

relative molecular mass of 10,387 ( $M_r$  10K), and a hydrophobic segment at the amino terminus with features characteristic of cleavable signal sequences<sup>10</sup>. This region of the protein also contains sequence elements (typically Leu-Leu-Ala-Gly-Cys) found in bacterial lipoproteins<sup>11</sup>. A search of the NBRF/PIR and SWISS-PROT databases using the FASTP program<sup>12</sup> did not detect any proteins with significant homology to *bor*.



FIG. 1 DNA and deduced amino-acid sequences of *lom* (b) and *bor* (a) (ref. 9) showing *TnphoA* fusion joints. The fusions identify the *bor* reading frame, which extends from an ATG codon at nucleotide 46,752 preceded by a possible ribosome binding site and several in-frame termination codons, to a TAA codon at nucleotide 46,461. Because *bor* fusion subclones in both orientations in the plasmid vector still have high levels of alkaline phosphatase activity, it is possible, though not certain, that the *bor* promoter lies to the left of the  $\lambda$  *BstY1* site at nucleotide 47,773. Fusion joints are marked above the sequence. Numbering is that of ref. 9. METHODS. To generate gene fusions,  $\lambda$ 16-25 was used to lysogenize JJ35 (*galK hsdR lacY metB supE supF tonA trpR* F'42::TnphoA), which was then induced, and the resulting lysate used in turn to lysogenize MPh2 (*araD*  $\Delta$ (*brnQ-phoR*)  $\Delta$ *lacU169 rpsL*). Phage that had transduced active *TnphoA* fusions were identified by selecting lysogens on plates containing kanamycin and the chromogenic alkaline phosphatase substrate XP. These transducing phage were purified for analysis. a, The *bor* region; DNA from phage 11.9 and 11.10 was digested with *BstY1* and the 4.5-kilobase (kb) *bor-phoA* fusion fragments were subcloned into *Bam*H1-digested pZ150 and sequenced as described<sup>29</sup>. b, The *lom* region; 5.5 kb *EcoRV-SalI* fusion fragments from phage 11.15 and 12.21 were subcloned into *EcoRV/SalI*-digested pZ150 and sequenced.

However, a small region of the conjugative plasmid ColV2-K94, involved in resistance to killing by serum and in surface exclusion, is highly homologous to this region of  $\lambda$  (ref. 5). The plasmid locus (*iss*) has homology to part of  $R_z$  and the sequences 3' to it, which encompass an open reading frame which could encode a protein with 79% overall identity to the *bor* gene product, including a frameshifted 10-amino-acid stretch of non-identity towards the middle of the protein (identity at the DNA level is 93%). It is possible that this locus originated in a fairly recent recombination event between ColV2-K94 and  $\lambda$ .

Fusions to *lom* (Fig. 1b) reveal an export signal in the first 71-amino-acid residues, consistent with the localization of the gene product to the outer membrane<sup>2</sup>.

The immunoprecipitation experiment shown in Fig. 2 demonstrates that fusion proteins of the expected sizes are made in the corresponding lysogens, and are present in amounts that accord with the observed alkaline phosphatase levels. These levels suggest, by comparison with fusions to other previously quantitated proteins<sup>13</sup>, amounts of ~100 molecules per cell for *lom* and 2,000 molecules per cell for *bor*.

Lysogeny by phage  $\lambda$  can increase the survival of some *E. coli* strains in serum<sup>14</sup>. This finding, and the striking homology between *bor* and the ColV2-K94 *iss* locus, prompted us to test whether *bor* and *lom* are involved in this phenomenon. Accordingly, *E. coli* strains lysogenic for  $\lambda$ +,  $\lambda$ 16-25 or its *lom::TnphoA* or *bor::TnphoA* derivatives were tested for their susceptibility to being killed by serum. The K-12 strain AB1157 was used in these experiments (see Table 1). The results show that lysogeny by either  $\lambda$  or  $\lambda$ 16-25 confers a >20-fold protection on AB1157, and that this protection is completely and specifically abolished by *TnphoA* insertions in *bor*. This effect on survival is not due to the presence of *TnphoA* because the *lom::TnphoA* insertion still allows protection. The data also suggest that *lom* does not contribute significantly to the observed effect, and that the transducing phage  $\lambda$ 16-25 provides a comparable degree of protection to  $\lambda$ +. It is also noteworthy that the degree of protection seen here is comparable to the roughly 20-fold effect reported for *iss* under similar conditions<sup>5</sup>.

These findings afford a new perspective on lysogeny by temperate bacteriophage. In particular, they demonstrate a novel type of selective advantage associated with the expression of prophage-encoded genes. This may turn out to be a fairly general phenomenon. For example, lysogenic conversion by certain phage is known to alter the cell-surface O antigen of *Salmonella typhimurium*<sup>15</sup>, and *E. coli* phage are known to produce porin-like outer-membrane proteins during lysogeny<sup>16</sup>. Phage-associ-

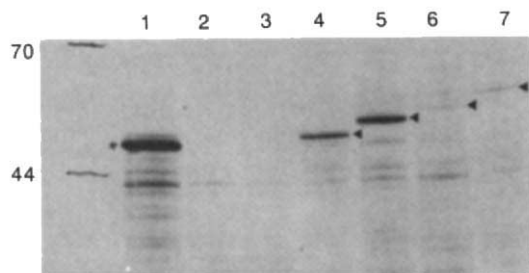


FIG. 2 Immunoprecipitation of alkaline phosphatase fusion proteins. Lanes: 1, XPh5 (*lac phoR rpsL*, a strain constitutive for wild-type alkaline phosphatase, with about 600 units of activity); 2, MPh2; 3, MPh2 ( $\lambda$ 16-25); 4, MPh2 ( $\lambda$ 16-25 *bor::TnphoA* 11.9); 5, MPh2 ( $\lambda$ 16-25 *bor::TnphoA* 11.10); 6, MPh2 ( $\lambda$ 16-25 *lom::TnphoA* 11.15); 7, MPh2 ( $\lambda$ 16-25 *lom::TnphoA* 12.21). Arrowheads indicate fusion proteins. Asterisk indicates wild-type alkaline phosphatase. *M<sub>r</sub>* standards (K) on left.

**METHODS.** Strains lysogenic for phage carrying fusions were grown to exponential phase in M63 glucose medium, pulse-labelled for 1 min with [<sup>35</sup>S]methionine, chased for 1 min with unlabelled methionine, and immunoprecipitated with anti-alkaline phosphatase antibody as described<sup>29</sup>. Labelled proteins were resolved on a 10% polyacrylamide gel.

TABLE 1 Killing of AB1157 and derivatives by 20% guinea-pig serum

| Strain                                             | % Survival       | <i>n</i> | <i>R</i> |
|----------------------------------------------------|------------------|----------|----------|
| AB1157                                             | 0.0013 (0.00041) | 16       | —        |
| AB1157 ( $\lambda$ +)                              | 0.030 (0.0050)   | 13       | 23.1     |
| AB1157 ( $\lambda$ 16-25)                          | 0.031 (0.0048)   | 21       | 23.9     |
| AB1157 ( $\lambda$ 16-25 <i>bor::TnphoA</i> 11.10) | 0.0011 (0.00026) | 22       | 0.9      |
| AB1157 ( $\lambda$ 16-25 <i>lom::TnphoA</i> 11.15) | 0.027 (0.0063)   | 13       | 20.1     |

AB1157 is *ara argE*  $\Delta$ (*gpt-proA*)62 *galK* *hisG* *kdgK* *lacY1* *leuB6* *mgI51* *mtl-1* *qsr*<sup>+</sup> *rac*<sup>+</sup> *rfaD1* *rpsL* *supE* *thi-1* *thr-1* *tsx* *xyl-5*. An isogenic derivative of this strain has been used to study serum protection mediated by *iss*<sup>5</sup>, and it lacks two cryptic lambdoid prophage<sup>30</sup> which may express proteins related to *lom* and *bor*. Mean survival values are shown for each strain with s.e.m. in parentheses. *n*, Number of replicate assays for each strain. *R*, mean % survival of each strain divided by that of AB1157. The protocol was essentially the same as that used in similar experiments involving *iss*<sup>5</sup>. Strains were grown overnight at 37 °C in LB medium, diluted 100-fold and grown for 105 min. Cells were pelleted and resuspended in PBS at  $\sim 2 \times 10^8$  ml<sup>-1</sup>, adjusted to equal concentrations, and diluted 1:10 into either prewarmed PBS or a 20% final concentration of guinea-pig complement (Gibco) in PBS (sterilized by filtration through a Costar 0.22  $\mu$ m cellulose acetate filter). Cells were incubated at 37 °C for 60 min with aeration, and dilutions plated for viable counts. Experiments were typically done in sextuplicate and repeated several times on different days. In some cases assays were done as mixtures of *phoA*<sup>+</sup> and *phoA*<sup>-</sup> strains, with the same result.

ated ampicillin resistance has also been reported<sup>17</sup>. *E. coli*  $\lambda$  lysogens (as well as lysogens of phage Mu, P1 and P2) have a significant growth advantage over nonlysogens under certain conditions, and this 'increased fitness' correlates with an altered profile of outer membrane proteins<sup>18-20</sup>. A DNA fragment carrying *cos* (which is adjacent to *bor*) can produce a similar effect<sup>21</sup>. Our results demonstrate the involvement of a specific phage gene in a selectively advantageous phenotype which may also be a significant virulence factor. Although temperate phage such as corynebacteriophage  $\beta$  have been implicated in disease by producing exotoxins<sup>22</sup>, this is the first example known to us describing a phage-encoded function that could provide direct protection from host immune defences. Resistance to serum-complement killing is one of the most extensively studied of bacterial virulence determinants<sup>3,4,23,24</sup>. Significantly, a subcloned fragment carrying *iss* confers a 100-fold increase in virulence on *E. coli* in chicks<sup>6</sup>. Whether  $\lambda$  itself can produce this effect is not known.

Recent observations suggest the possibility that *lom* may also affect the properties of the *E. coli* lysogen in an animal host: this gene displays significant homology to virulence loci involved both in macrophage survival (in *Salmonella typhimurium*) and invasiveness (in *Yersinia enterocolitica*) (W. S. Pulkkinen, V. L. Miller and S. I. Miller, manuscript submitted; and ref. 25). The conservation of *lom* and *bor* sequences among other phage and in clinical bacterial isolates will be of interest in the context of these observations. For example, the *E. coli* shiga-like toxin-converting phage H-19B has homology to both the *J-lom* and *P-Q* regions of  $\lambda$  (ref. 26), and the lambdoid phage P22 and  $\Phi$ 80 encode lysogenic conversion functions that map in regions corresponding to *lom* (refs 15, 27). Furthermore, sequences related to *iss*—and therefore to *bor*—have been detected in 95.5% of ColV<sup>+</sup> *E. coli* strains isolated from the blood of patients with bacteraemia, from 68.8% of those with intestinal infections, and from about 28% of all clinical strains examined, regardless of ColV production<sup>28</sup>. The possibility is raised that these sequences will be widespread. □

Received 7 March; accepted 7 June 1990.

1. Court, D. & Oppenheim, A. B. in *Lambda II* (eds Hendrix, R., Roberts, J. W., Stahl, F. W. & Weisberg, R. A.) 251-277 (Cold Spring Harbor Laboratory, New York, 1983).
2. Reeve, J. N. & Shaw, J. E. *Molec. gen. Genet.* **172**, 243-248 (1979).
3. Taylor, P. *Microbiol. Rev.* **47**, 46-83 (1983).
4. Joiner, K. A. *Rev. Microbiol.* **42**, 201-230 (1988).
5. Chuba, P., Leon, M., Banerjee, A. & Palchoudhuri, S. *Molec. gen. Genet.* **216**, 287-292 (1989).



6. Binns, M. M., Davies, D. & Hardy, K. *Nature* **279**, 778-781 (1979).
7. Lutkenhaus, J., Wolf-Watz, H. & Donachie, W. D. *J. Bact.* **142**, 615-620 (1980).
8. Manoil, C., Mekalanos, J. & Beckwith, J. *J. Bact.* **172**, 515-518 (1990).
9. Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. & Petersen, G. B. *J. molec. Biol.* **162**, 729-773 (1982).
10. Von Heijne, G. *Nucleic Acids Res.* **14**, 4683-4690 (1986).
11. Perumal, N. & Minkley, E. G. *J. biol. Chem.* **259**, 5357-5360 (1984).
12. Pearson, W. R. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2444-2448 (1988).
13. Shuman, H. A. *et al.* *J. biol. Chem.* **255**, 168-174 (1980).
14. Muschel, L. H., Ahl, L. A. & Baron, L. S. *J. Bact.* **96**, 1912-1914 (1968).
15. Poteete, A. R. in *The Bacteriophages* (ed. Calendar, R.) 647-682 (Plenum, New York, 1988).
16. Blasband, A. J., Marcotte, W. R. & Schnaitman, C. A. *J. biol. Chem.* **261**, 12723-12732 (1986).
17. Smith, H. W. *Nature New Biol.* **238**, 205-206 (1972).
18. Edlin, G. *Nature* **255**, 735-737 (1975).
19. Edlin, G., Lin, L. & Bitner, R. *J. Virol.* **21**, 560-564 (1977).
20. Edlin, G. in *Persistent Viruses* (eds Stevens, J. G., Todaro, G. J. & Fox, C. F.) 1-14 (Academic, New York, 1978).
21. Edlin, G., Tait, R. & Rodriguez, R. *Bio/technology* **2**, 251-254 (1984).
22. Bishai, W. R. & Murphy, J. R. in *The Bacteriophages* (ed. Calendar, R.) 683-724 (Plenum, New York, 1988).
23. Brubaker, R. R. *A. Rev. Microbiol.* **39**, 21-50 (1985).
24. Timmis, K. N., Boulnois, G. J., Bitter-Suermann, D. & Cabello, F. C. *Curr. Top. Microbiol. Immunol.* **118**, 197-218 (1985).
25. Miller, V. L., Bliska, J. B. & Falkow, S. *J. Bact.* **172**, 1062-1069 (1990).
26. Huang, A., Friesen, J. & Brunton, J. *J. Bact.* **169**, 4308-4312 (1987).
27. Rybchin, V. N. *Gene* **27**, 3-11 (1984).
28. Fernandez-Beros, M. E., Kissel, V., Lior, H. & Cabello, F. C. *J. clin. Microbiol.* **28**, 742-746 (1990).
29. Boyd, D., Manoil, C. & Beckwith, J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8525-8529 (1987).
30. Anilionis, A., Ostapchuk, P. & Riley, M. *Molec. gen. Genet.* **180**, 479-481 (1980).

ACKNOWLEDGEMENTS. We thank D. Court, D. Boyd and J. Mekalanos for helpful discussions, S. Miller and V. Miller for sharing unpublished results, and members of our laboratory for comments on the manuscript. This work was supported by the American Cancer Society (J.B.) and the Albert J. Ryan Foundation (J.J.B.).

## CORRECTION

### Natural occurrence of silicon carbide in a diamondiferous kimberlite from Fuxian

Irene Leung, Wenxiang Guo, Irving Friedman & Jim Gleason

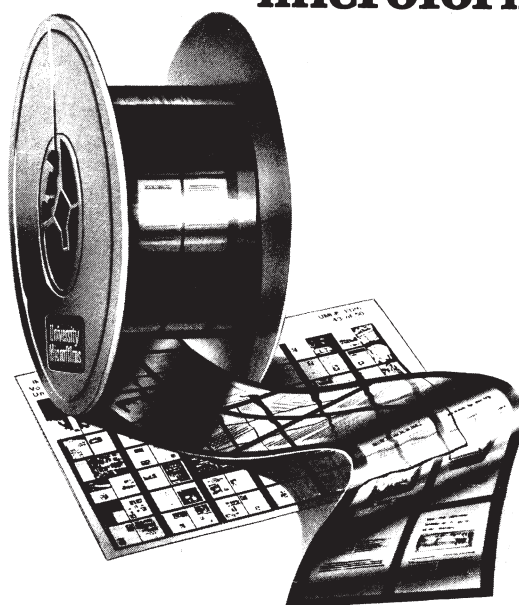
*Nature* **346**, 352-354 (1990).

THIS letter as printed in the 26 July issue contained some errors in the last two paragraphs (page 354) such that in places carbon-13 should read carbon-12 and vice versa. The section affected by this error is reproduced here:

Isotopically zoned but non-coated diamonds<sup>23,26</sup> show variations in <sup>13</sup>C from core to rim of only about 4‰. We suggest that the 20‰ fractionation observed between SiC and diamond is due to a distillation process: preferential uptake of <sup>12</sup>C during the growth of SiC enriched the carbon reservoir, in which the diamond later grew, in <sup>13</sup>C. Evidence that the SiC was formed first is provided by the presence of SiC inclusions in diamonds from Fuxian<sup>13</sup>. To explain the preferential uptake of <sup>12</sup>C by SiC, we examined the thermodynamic properties (1 atm) of diamond<sup>27</sup> and 6H SiC<sup>28</sup> and found that SiC has much higher enthalpy and entropy than diamond. Deines<sup>29</sup> noted that in an isotope exchange reaction, the solid of lower heat capacity can be correlated with higher vibrational frequency and a stronger tendency to incorporate the heavy isotope. This suggests that SiC should be enriched in <sup>12</sup>C and diamond in <sup>13</sup>C, as we observe.

In studies of iron meteorites<sup>30,31</sup>, systematic <sup>13</sup>C enrichment in graphite and <sup>13</sup>C depletion in cohenite (iron carbide), in 12‰ fractionation at ~600 °C was observed, and the fractionation was expected to increase with increasing temperature. The <sup>13</sup>C-<sup>12</sup>C systematics observed in our SiC-diamond system is completely analogous. Our isotope data imply that the carbon reservoir in the mantle beneath Fuxian has a  $\delta^{13}\text{C}$  value between -24‰ and -4‰, and is probably much lower than -5‰, a value derived from isotope analysis of natural diamonds. A more precise estimate would require a knowledge of all fractionation processes and the number of carbon-bearing phases participating in the isotope exchange reaction. □

nature  
is available in  
microform.



University Microfilms

International reproduces this publication in microform: microfiche and 16mm or 35mm film. For information about this publication or any of the more than 13,000 titles we offer, complete and mail the coupon to: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106. Call us toll-free for an immediate response: 800-521-3044. Or call collect in Michigan, Alaska and Hawaii: 313-761-4700.

☐ Please send information about these titles:

\_\_\_\_\_

Name \_\_\_\_\_

Company/Institution \_\_\_\_\_

Address \_\_\_\_\_

City \_\_\_\_\_

State \_\_\_\_\_ Zip \_\_\_\_\_

Phone ( ) \_\_\_\_\_

University  
Microfilms  
International